OPTIMIZATION BY GAUSSIAN SMOOTHING WITH APPLICATION TO GEOMETRIC ALIGNMENT

BY

HOSSEIN MOBAHI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Computer Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor Yi Ma, Chair and Director of Research
Professor Thomas S. Huang
Professor David A. Forsyth
Professor Derek W. Hoiem
Professor Stefano Soatto, UCLA

ABSTRACT

It is well-known that global optimization of a nonconvex function, in general, is computationally intractable. Nevertheless, many objective functions that we need to optimize may be nonconvex. In practice, when working with such a nonconvex function, a very natural heuristic is to employ a coarse-to-fine search for the global optimum. A popular ***deterministic*** procedure that exemplifies this idea can be summarized briefly as follows. Consider an unconstrained optimization task of minimizing some nonconvex function. One starts from a highly ***"smoothed"*** version of the objective function and hopes that the smoothing eliminates most spurious local minima. More ideally, one hopes that the highly smoothed function would be a convex function, whose global minimum can be found efficiently. Once the minimum of the smoothed function is found, one could gradually reduce the smoothing effect and follow the continuous path of the minimizer, eventually towards a minimum of the objective function. Empirically, people have observed that the minimum found this way has high chance to be the global minimum.

Despite its empirical success, there has been little theoretical understanding about the effect of smoothing on optimization. This work rigorously studies some of the fundamental properties of the smoothing technique. In particular, we present a formal definition for the functions that can eventually become convex by smoothing. We present extremely simple sufficient condition for asymptotic convexity as well as a very simple form for an asymptotic minimizer. Our sufficient conditions hold when the objective function satisfies certain decay conditions.

Our initial interest for studying this topic arise from its well-known use in geometric image alignment. The alignment problem can be formulated as an optimization task that minimizes the visual difference between the images by searching the space of transformations. Unfortunately, the cost function associated to this problem usually contains many local minima. Thus, unless

very good initialization is provided, simple greedy optimization may lead to poor results.

To improve the attained solution for the alignment task, we propose smoothing the objective function of the alignment task. In particular, we derive the theoretically correct image blur kernels that arise from (Gaussian) smoothing an alignment objective function. We show that, for smoothing the objective of common motion models, such as affine and homography, there exists a corresponding ***integral operator on the image space*** . We refer to the kernels of such integral operators as ***transformation kernels*** . Thus, instead of convolving the objective function with a Gaussian kernel in transformation space, we can equivalently compute an integral transform in the image space, which is much cheaper to compute.

To my Absolute Value
Mahsa
for her Unbounded Support

To the New Dimension
Artin
for becoming my Inflection Point

And to my Real Roots
Khalil & Maghbouleh
for their Existence and Uniqueness

ACKNOWLEDGMENTS

I will forever be in debt to my PhD adviser, Yi Ma, for being such a peerless mentor in all aspects. I am extremely thankful for his priceless advice, encouragement, patience and understanding throughout the journey of my PhD. Yi taught me an exceptionally deep and rigorous treatment of computer vision that fundamentally has changed my definition of scientific research in this area. Yi, thank you for your emotional support during my rough days. You have been a perfect adviser in every sense and you made my PhD the most pleasant academic experience.

I would also like to thank my PhD committee members, Thomas S. Huang, David A. Forsyth, Derek W. Hoiem, and Stefano Soatto for their invaluable guidance and thoughtful comments. With your deep insight, you have provided a wealth of connections and extensions that I could explore, and you have shaped a list of interesting research problems that I would like to explore after graduation.

My PhD adventure began in inner turmoil. Throughout the first two years, I found myself often lost and stressed. However, with the help of caring faculty, I regained my confidence and thrived. Thank you for your endless kindness and support, Mehdi T. Harandi, Thomas S. Huang, Karrie G. Karahalios, and Klara Nahrstedt. As well, you will always be remembered Sylvian R. Ray for your incredible trust and support. Thank you my peers, Tony Bergstrom and Jacob Schroder, for keeping me away from the stress. You remained my closest friends in school until graduation.

The hardships of my first few years were ameliorated upon meeting Yi Ma and several other distinguished researchers. My academic life flourished after my contact with them, either through internships or informally. I would like to express my deep gratitudes to Alexis Guigue (UBC), Nebojsa Jojic (MSR), Ce Liu (MSR), Alessandro Perina (MSR), Stephen Montgomery-Smith (Mizzou), Jason Weston (Google Research), Vadim Zharnistky (Math@UIUC),

and Larry Zitnick (MSR).

Around the same time, I discovered the incredible value of discussing research ideas with my peers. The closest circle was my smart and hard working lab mates. They raised the standards for being a PhD student way beyond my imagination. It was difficult not to move forward when being among them. I owe you a big chunk of everything I learned during these years; for all the technical discussions, even after you graduated. Thank you Arvind Ganesh, Shankar R. Rao, Zihan Zhou, Drew Wagner, John Wright, Allen Y. Yang, and Guangcan Liu. The next layer of the onion were my bright and clever peers in the computer science department. I was very lucky to have these folks close to me as they were always open to discussing various problems and questions that arose, and I have learned a great deal through these discussions. Thanks for always being there despite being busy with your own deadlines and homework, Yonatan Bisk, Mingwei Chang, Ian Endres, Yuntao Jia, Miles Johnson, Kevin Karsch, Nicolas Loeff, Victor Lu, Amir Nayyeri, Alexander Sorokin, and Dmitry Yershov.

The last two years of my PhD were when most of the ideas in this thesis took shape. Throughout this period, I have been financially supported by UIUC's CSE fellowship. I would like to thank the CSE office, and Michael Heath in particular.

We now get to a point where time, space, and words are impotent to capture their value. My soul mate Mahsa and my adorable six month son Artin have been the ultimate source of love and hope throughout difficult times. Mahsa, you carried the heavy weight of life's challenges on shoulder throughout these years so that I could concentrate and be productive at school. Without your unparalleled support, encouragement, and sacrifices, my PhD journey would not have come to an end. I would also like to thank my sister Hedyeh, and Mahsa's mom, Mahrokh, for being here with Artin and creating a comfortable and soothing home for me. Last but not least, I would like to thank my parents, Khalil and Maghbolueh, for years of unconditional love and incredible sacrifices.

TABLE OF CONTENTS

LIST OF FIGURES ix

NOTATIONS AND DEFINITIONS xi

CHAPTER 1 INTRODUCTION 1
1.1 Nonconvex Optimization 1
1.2 Homotopy Continuation 3
1.3 Smoothing 4
1.4 Contributions 7

CHAPTER 2 ASYMPTOTIC ANALYSIS 9
2.1 Assumptions and Notations 10
2.2 Definitions 10
2.3 Basic Properties 11
2.4 Derivative-Free Results for Decaying Functions 13
2.5 The Function Space over $\mathbb{R}^n$ 21
2.6 Proofs 23

CHAPTER 3 TRANSFORMATION KERNELS 56
3.1 Motivation 57
3.2 Notation and Definitions 59
3.3 Smoothing the Objective 60
3.4 Transformation Kernels 62
3.5 Computation of the Integral Transform 66
3.6 Regularization 68
3.7 Proofs 69

CHAPTER 4 2D APPLICATIONS 82
4.1 Computation with Homography Kernel 82
4.2 Quantitative Results on Planar Scenes 87
4.3 Qualitative Results on 3D Reconstruction 88

CHAPTER 5 3D APPLICATIONS 94
5.1 Problem Formulation 94
5.2 Smoothing 96
5.3 Gradients 97
5.4 Asymptotic Properties 97
5.5 Illustrative Example 97
5.6 Algorithm & Results 100
5.7 Proofs 101

CHAPTER 6 CONCLUSION & FUTURE DIRECTIONS 108
6.1 Continuation 108
6.2 Kernels 109

REFERENCES 111

LIST OF FIGURES

1.1 Time evolution of a Gaussian function under heat (top) and Schrodinger (bottom) equations. Time progression is from left to right. 5

2.1 **Left to Right**: The function $g(x;\sigma) = [f \star k(\,.\,;\sigma^2)](x)$, where $f(x) = e^{-\frac{x^2}{2\epsilon^2}} - e^{-\frac{x^2\epsilon^2}{2}}$, with increasing values of σ. Nonconvex regions are colored by pink. 11

2.2 $g(x;\sigma) = f(x) \star k(x;\sigma), f(x) = e^{-\frac{(x-1)^2}{0.1}} - e^{-x^2}$ for different choices of σ. Notice that the minimizer becomes unique for large σ and approaches the yellow line $x = -0.46247$. . . . 18

2.3 The taxonomy of the space of functions $\{f : \mathbb{R}^n \to \mathbb{R}\}$. 22

3.1 Basin of attraction for scale alignment. Egg shape input images are shown in (a) and (b), where black and white pixels are respectively by -1 and 1 intensity values. Obviously, the correct alignment is attained at $\theta = -1$, due to reflection symmetry. The objective function for z_{LK} is shown in (c) and for z in (d). Blue, green and red respectively indicate local maxima, global maximum and basin of attraction originating from local maxima of highest blur. . . 60

3.2 Visualization of affine and homography kernels specified by $\boldsymbol{A}_0 = [2 \quad 0.2\,; -0.3 \quad 4], \boldsymbol{b}_0 = [0.15 \quad -0.25]$ (also $\boldsymbol{c}_0 = [1 \quad -5]$ for homography). Here $\boldsymbol{x} \in [-1,1] \times [-1,1]$, $\sigma = 0.5$ and $\boldsymbol{y} = (1,1)$ or $\boldsymbol{y} = (0,0)$. More precisely, affine kernels in (a) $u(\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{x}, \boldsymbol{y} = (0,0))$ (b) $u(\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{x}, \boldsymbol{y} = (1,1))$ and homography kernels in (c) $u(\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{x}, \boldsymbol{y} = (0,0))$ (d) $u(\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{x}, \boldsymbol{y} = (1,1))$ 62

4.1 Dataset provided by [1] consisting of five planar scenes, each having six different views of increasingly dramatic perspective effect. 87

4.2 **Top**: Representative rectified views from the dataset provided in [1]. **Bottom**: NCC value after alignment. Horizontal axis is the view index (increasing in complexity) of the scene. Four views are used for each scene, each one being as f_2 and compared against f_1, which is a rectified view in the dataset. 89
4.3 A pair of matched regions from the same facade with different partial occlusion. 90
4.4 Frontal and top views of the recovered building. Each pyramid shows the estimated location of a camera. 92
4.5 (a) Segmented and rectified facade. (b),(c) Same task from a different view. (c) Segmentation result refined to the orange box by matching. (d) Point-wise match between two regions of the facades using our method. (e) Feature-point matching result of the two rectified regions by SIFT [2], with red lines indicating mismatches. 93

5.1 Each point in $\mathcal{Q}$ is put into correspondence with one of the points in $\mathcal{P}$. 95
5.2 Optimization landscape for minimizing the function (5.19). The spectrum from blue to red indicates small to large values. 98
5.3 Top Row : Input $\mathcal{P}$, which is a rotated version of $\mathcal{Q}$. Middle Row : Transformed $\mathcal{P}$ to match $\mathcal{Q}$ using ICP. Bottom Row: Transformed $\mathcal{P}$ to match $\mathcal{Q}$ using proposed method. 105
5.4 Top Row : Input $\mathcal{P}$, which is a rotated version of $\mathcal{Q}$. Middle Row : Transformed $\mathcal{P}$ to match $\mathcal{Q}$ using ICP. Bottom Row: Transformed $\mathcal{P}$ to match $\mathcal{Q}$ using proposed method. 106
5.5 Top Row : Input $\mathcal{P}$, which is a rotated version of $\mathcal{Q}$. Middle Row : Transformed $\mathcal{P}$ to match $\mathcal{Q}$ using ICP. Bottom Row: Transformed $\mathcal{P}$ to match $\mathcal{Q}$ using proposed method. 107

NOTATIONS AND DEFINITIONS

$\triangleq$		Means equal by definition
x	$\triangleq$	Scalar
$\boldsymbol{x}$	$\triangleq$	Vector
$\boldsymbol{X}$	$\triangleq$	Matrix
$\mathcal{X}$	$\triangleq$	Set
$f(.)$	$\triangleq$	Scalar Valued Function
$\boldsymbol{f}(.)$	$\triangleq$	Vector Valued Function
$f_i(.)$	$\triangleq$	The i'th Component of Vector Valued Function
$\boldsymbol{I}$	$\triangleq$	The Identity Matrix
$\boldsymbol{X} \succeq \boldsymbol{Y}$	$\equiv$	$\forall \boldsymbol{v}\ ;\ \boldsymbol{v}^T(\boldsymbol{X}-\boldsymbol{Y})\boldsymbol{v} \geq 0$
$\|\boldsymbol{x}\|$	$\triangleq$	$\|\boldsymbol{x}\|_2$
$\|f\|$	$\triangleq$	$\sqrt{\int_{\mathcal{X}} f^2(\boldsymbol{x})\, d\boldsymbol{x}}$
$\mathbb{R}^n_+$	$\triangleq$	$\{x \in \mathbb{R}^n \,\|\, x \geq 0\}$
$\mathbb{R}^n_{++}$	$\triangleq$	$\{x \in \mathbb{R}^n \,\|\, x > 0\}$
$\mathcal{B}^n(\boldsymbol{x}_0, \rho)$	$\triangleq$	$\{\, \boldsymbol{x} \,\|\, \boldsymbol{x} \in \mathbb{R}^n \, \|\boldsymbol{x}_0 - \boldsymbol{x}\| < \rho \,\}$
∇	$\triangleq$	$\nabla_{\boldsymbol{x}}$
$\dot{\boldsymbol{x}}$	$\triangleq$	$\dfrac{\partial}{\partial \sigma}\boldsymbol{x}$
Δ	$\triangleq$	$\displaystyle\sum_{d=1}^{n} \frac{\partial^2}{\partial x_d^2}$
$\dfrac{d}{dx} f(u(x))$	$\triangleq$	$f'\big(u(x)\big)$
$\dfrac{d}{dx}\Big(f(u(x))\Big)$	$\triangleq$	$f'(u(x))u'(x)$
$k(\boldsymbol{x}; \sigma^2)$	$\triangleq$	$(2\pi\sigma^2)^{-\frac{\dim(\boldsymbol{x})}{2}} e^{-\frac{\|\boldsymbol{x}\|^2}{2\sigma^2}}$
$K(\boldsymbol{x}; \boldsymbol{\Sigma})$	$\triangleq$	$(2\pi \det(\boldsymbol{\Sigma}))^{-\frac{\dim(\boldsymbol{x})}{2}} e^{-\frac{\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}}{2}}$

CHAPTER 1

INTRODUCTION

It is well-known that global optimization of a nonconvex function, in general, is computationally intractable [3]. Unfortunately, many objective functions that arise in real applications are nonconvex. Good news is, however, real problems often have some kind of regularity and structure. Sometimes, by recognizing and exploiting these structures, it is possible to find a reasonable solution for a non-convex optimization task in reasonable time. In this dissertation, we focus on a particular structure, in which local minima are brittle and easy to filter out by smoothing. We provide intuitive and formal definitions for the smoothing concept. We describe how smoothing can be utilized within a continuation framework. The latter essentially means following the path of the minimizer of the smoothed function back to the original objective.

1.1 Nonconvex Optimization

Nonconvex optimization methods can be broadly categorized into ***deterministic*** and ***stochastic*** methods; regardless of the ability to finding the global minimum or just a local minimum. As their names suggest, the deterministic methods generate the same solution every time they run. However, the stochastic methods may find completely different solution every time they are used.

A representative example of the stochastic methods is ***simulated annealing*** (SA) [4]. This algorithm is inspired from annealing in metallurgy which involves heating and controlled cooling of materials. SA algorithm is an iterative scheme, where each iteration replaces the current solution with a random one. The distribution by which new samples are drawn is such that the points closer to the current solution are more likely to be chosen. The

newly sampled point is accepted as a solution with a probability that depends on the improvement in value function as well as a temperature parameter T. When T is large, the previous and new solutions are almost independent, and as T gets smaller, new solutions with improved value are preferred. The algorithm starts from a large value of T, which may help escape from local minima, and then continue to smaller values of T. SA can converge to global minimizer almost surely [4], but often with an extremely low convergence rate.

Among the deterministic techniques, ***gradient descent*** is perhaps the simplest and by far the most well-known method. This is a first-order iterative algorithm. In each iteration, the method takes a small step in the opposite direction of the gradient, hence the name descent. The algorithm is supposed to stop at a local minimum, because there exists no descent direction in that vicinity. While effective for convex tasks or nonconvex problems with equally good local minima, it often leads to unsatisfactory solution in a typical nonconvex task.

In order to find reasonable solution in reasonable time for nonconvex optimization tasks, it is often necessary to recognize and exploit some nice structure in the problem. There exists a vast literature in nonconvex optimization for handling different kinds of structure. For example, if a polynomial optimization can be closely approximated by ***sum of squares*** (SOS), then one can use a hierarchy of SOS relaxations, each of which can be formulated as a semidefinite program [5]. Another example is when the objective function can be written as ***difference of convex*** (DC) functions $f_1(\boldsymbol{x}) - f_2(\boldsymbol{x})$. The combination is no longer convex, but by alternating between fixing one and optimizing the other, one switches between convex and concave programs. While the convex part $f_1(\boldsymbol{x})$ can be efficiently solved, the concave part $-f_2(\boldsymbol{x})$ can be replaced by its convex envelope. When the domain of $\boldsymbol{x}$ is a ***simplex*** , convex envelope of a concave function has the simple form of an ***affine*** function [6], hence again easy to minimize.

Another method for nonconvex optimization, which is the central topic of this dissertation, is ***continuation*** or ***homotopy continuation*** method. The idea is to somehow simplify the original optimization task, e.g. by some convex relaxation, and then continuously deform it back to the original objective. While this deformation is happening, one follows the minimizer of the simplified problem back to the original problem. Some related topics

around this method will be reviewed in the rest of this chapter.

1.2 Homotopy Continuation

Continuation is a well established procedure in numerical analysis. The main idea is to smoothly deform an easy problem into the actual problem, while tracking the solution of the deformed problems. The function which describes this deformation is called a ***homotopy map*** . Suppose we would like to solve a problem that involves a function $\boldsymbol{f} : \mathcal{X} \to \mathcal{Y}$. Homotopy continuation method embeds $\boldsymbol{f}(\boldsymbol{x})$ into a parameterized family of functions $\boldsymbol{g}(\boldsymbol{x}, t)$, where $\boldsymbol{g} : \mathcal{X} \times \mathcal{T} \to \mathcal{Y}$ and $\mathcal{T} = \{\, t \,|\, 0 \leq t \leq 1\}$. The embedding should have the property that $\boldsymbol{g}(\boldsymbol{x}, 0)$ is easy to solve and $\boldsymbol{g}(\boldsymbol{x}, 1) = \boldsymbol{f}(\boldsymbol{x})$.

For example, suppose the goal is to solve a system of nonlinear polynomial equations $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$. Assume it is difficult to solve this system, but we know the solution of another system $\boldsymbol{f}^{\dagger}(\boldsymbol{x}) = \boldsymbol{0}$, which has the same number of variables and the same polynomial total degree as $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$. Consider a smooth homotopy map $\boldsymbol{g}$ such that $\boldsymbol{g}(\boldsymbol{x}, 0) = \boldsymbol{f}^{\dagger}(\boldsymbol{x})$ and $\boldsymbol{g}(\boldsymbol{x}, 1) = f(\boldsymbol{x})$. Then, under certain regularity conditions [7, 8], the roots of this system can be conceptually found by following the curve of solutions $\boldsymbol{x}(t)$ of the system $\boldsymbol{g}(\boldsymbol{x}, t) = \boldsymbol{0}$ while t changes from 0 to 1.

A very simple homotopy is the ***convex combination*** map $\boldsymbol{g}(\boldsymbol{x}, t) = t\boldsymbol{f}(\boldsymbol{x}) + (1 - t)\boldsymbol{f}^{\dagger}(\boldsymbol{x})$. This homotopy is very popular in the context of solving polynomial systems. In this dissertation, however, we focus on a more sophisticated homotopy that performs Gaussian smoothing, i.e. $g(\boldsymbol{x}, t) = [f \star k](\boldsymbol{x})$, where $\star$ is the convolution operator and k is an isotropic Gaussian kernel,

$$k \triangleq \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\|\boldsymbol{x}\|^2}{2\sigma^2}} . \tag{1.1}$$

Then, we can have the bandwidth σ of the Gaussian change with time t, for example via $t = \frac{1}{1+\sigma^2}$ for $\sigma > 0$ and $0 < t < 1$.

1.3 Smoothing

The notion of smoothing is quite intuitive. Loosely speaking, a smoothing process ***spreads*** some given mass across the domain. For example, if the original form has a sharp peak at a certain position, smoothing should decrease the mass at that location and increase the mass in its surrounding. Nevertheless, there is no unique rigorous definition for smoothing[1]. Here, we provide two example definitions that are consistent with the aforementioned intuition.

Smoothing can be formally described as a time evolution process. Starting from an original form $f(\boldsymbol{x})$, the smoothing ***gradually*** deforms $f(\boldsymbol{x})$ by spreading its mass over time. This viewpoint on smoothing provides an immediate framework for its rigorous treatment using ***partial differential equations*** (PDE).

$$\begin{aligned} \frac{\partial}{\partial t} g(\boldsymbol{x}, t) &= Dg(\boldsymbol{x}, t) && (1.2) \\ g(\boldsymbol{x}, 0) &= f(\boldsymbol{x}) && (1.3) \\ g(\boldsymbol{x}, t)_{|\boldsymbol{x} \in \partial\mathcal{X}} &= h(\boldsymbol{x}, t)\,, && (1.4) \end{aligned}$$

where D is a ***differential operator***, $f : \mathcal{X} \to \mathbb{R}$ is the initial condition and $g : \mathcal{X} \times \mathcal{T}$ represents the evolved smoothed form. In addition, if the domain $\mathcal{X}$ is bounded, $h : \partial\mathcal{X} \times \mathcal{T}$ specifies the value of g at the boundary.

For example, setting the differential operator D to Δ or $i\Delta$ results in the ***heat equation*** and ***Schrodinger's equation*** respectively. Here i is the imaginary unit and Δ denotes the ***Laplace operator*** w.r.t. spatial variable, i.e. the first argument of g. Observe that, the heat and Schrodinger's equations are only different in the imaginary factor. Thus, they are expected to behave somewhat similarly, except that the imaginary factor causes some ***oscillation*** in the solution of Schrodinger's equation. The smoothing effect of these operators is illustrated in Figure 1.1.

The heat kernel is of particular interest in this dissertation, due to its useful properties for nonconvex optimization. Specifically, it tends to damp high frequency components of f. In order to see that, it is simpler to look at the

[1]The smoothing in our context should not be confused with the technical term of smooth functions, which have the precise definition of being infinitely differentiable.

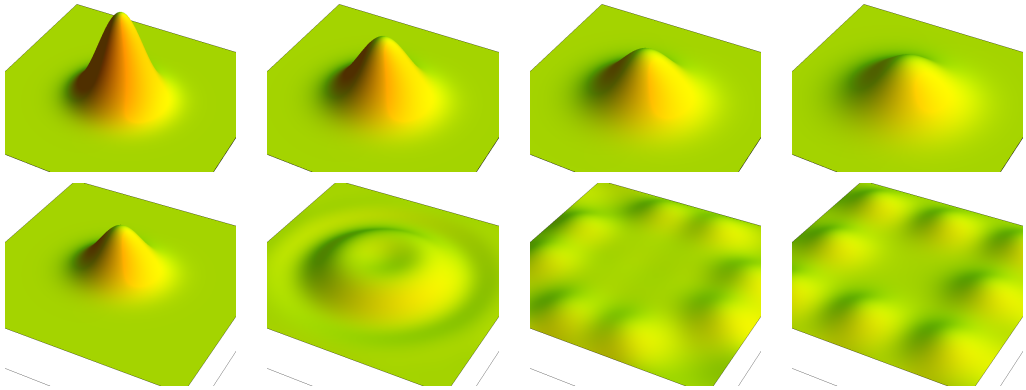


Figure 1.1: Time evolution of a Gaussian function under heat (top) and Schrodinger (bottom) equations. Time progression is from left to right.

solution of the heat equation when the domain is unbounded, i.e. $\mathcal{X} = \mathbb{R}^n$. It is known that, the following PDE,

$$\begin{aligned} \frac{\partial}{\partial \sigma} g(\boldsymbol{x}, \sigma) &= \sigma \Delta g(\boldsymbol{x}, \sigma) && (1.5) \\ g(\boldsymbol{x}, 0) &= f(\boldsymbol{x}) \,. && (1.6) \end{aligned}$$

has a solution of the following form,

$$g(\boldsymbol{x}, \sigma) = [f \star k_\sigma](\boldsymbol{x}) \,, \tag{1.7}$$

where $\star$ is the convolution operator (w.r.t. variable $\boldsymbol{x}$) and k_σ is the isotropic Gaussian kernel with variance σ^2, as defined in (1.1). Therefore the Fourier transform of $g(\boldsymbol{x}, \sigma)$ w.r.t. the spatial variable $\boldsymbol{x}$, can be expressed as,

$$\begin{aligned} \hat{g}(\boldsymbol{\omega}; \sigma) &= \hat{f}(\boldsymbol{\omega}) \hat{k}(\boldsymbol{\omega}; \sigma) && (1.8) \\ &= 2\pi\sigma \hat{f}(\boldsymbol{\omega}) k(\boldsymbol{\omega}; \frac{1}{\sigma}) \,. && (1.9) \end{aligned}$$

It is now apparent that high frequency components always get attenuated and more aggressively when σ is large.

This property of the heat kernel is of interest for nonconvex optimization, because it may suppress small fluctuations and thus eliminate brittle local

Algorithm 1 Minimization by Gaussian Smoothing Continuation.

1: Input: $f : \mathcal{X} \to \mathbb{R}$, The set $\{\sigma_k\}$ for $k = 1, \ldots, K$ s.t. $0 < \sigma_{k+1} < \sigma_k$.
2: **for** $k = 1 \to K$ **do**
3: $\quad \boldsymbol{x}_k =$ local minimizer of $g(\boldsymbol{x}; \sigma_k)$ initialized at $\boldsymbol{x}_{k-1}$
4: **end for**
5: Output: $\boldsymbol{x}_K$

minima on the surface of f. Thus, instead of optimizing f directly, we optimize g as a surrogate. There is a trade-off here though; while the higher σ is better at suppressing local minima, it also deforms f from its original form more significantly. Hence, the global minimizer of g may not necessarily coincide with even a local minimizer of f.

A heuristic for coping with this issue is applying continuation scheme. That is, starting from a highly smoothed objective, i.e. $g(\boldsymbol{x}; \sigma)$ with a large σ. Hopefully, for a large enough σ, the local minima of $f(\boldsymbol{x})$ disappear and $g(\boldsymbol{x}; \sigma)$ becomes convex. Therefore, it is easy to find its global minimizer. Then, we gradually deform g back to f by shrinking σ toward zero. At the same time, we follow the curve that this minimizer traces out through the deformation. The hope is that, one finds the global or at least a good local minimum of f using this heuristic. The algorithmic sketch of this idea is explained in Algorithm 1.

The underlying idea in Algorithm 1 has been widely utilized for optimization in different disciplines for a long history. Examples include graduated optimization [9], optimization by homotopy continuation [10], deterministic annealing [11], diffusion equation method [12], etc. In particular, in machine learning community, there has been an increasing interest in such concepts with applications to semi-supervised kernel machines [13], multiple instance learning [14,15], semi-supervised structured output [16], and statistical state estimation [17]. Interested reader is referred to the survey [18] (written from physicists perspective). Similar concept for discrete spaces has also been used [19–22].

While homotopy continuation for solving polynomial systems has a solid theory, continuation for optimization has remained at a heuristic level. This is mainly because the homotopy methods for solving polynomial systems are exhaustive, i.e. they enumerate all the roots of the system. So, as an optimization tool, they can be used to find ***all stationary points*** of the

objective, by zero crossing the gradient of the objective [23, 24]. However, this is very inefficient and often impractical, as the number of local minima may be large and likely to grow exponentially in dimension.

1.4 Contributions

The contribution of this thesis is two-fold. From theoretical perspective, we address some fundamental issues associated with optimization by Gaussian smoothing continuation.

Surprisingly, despite its long history, tremendous popularity, and deep connections to fundamental concepts in physics and mathematics, there has been little theoretical understanding about the effect of smoothing in optimization and the continuous path of the minimizers associated with the process of gradual smoothing. For example, this approach is the most useful for a function when smoothing can ***ultimately*** lead to a convex function. So a natural question is what conditions on f are required to guarantee this.

Even if we know that a given function can become convex by smoothing, there are still outstanding questions. For example, severe smoothing of the objective function makes the resulting function almost flat. This causes troubles for many numerical procedures in finding the minimizer of such (convex) functions; the solution can be extremely unstable and inaccurate. Thus, it is important to know if there is any closed-form solution for such a minimizer.

Answers to these questions can help us understand for what kind of functions we should expect the smoothing technique to work the most effectively. This dissertation presents a formal definition for the functions that can eventually become convex by smoothing. We refer to such functions as ***asymptotically convex functions*** . We show that, under mild conditions, asymptotically convex functions are a nontrivial superset of convex functions and inherit some of the nice properties of convex functions. For example, they obey some form of gradient inequality and positive-definite Hessian. We present an extremely simple and ***derivative-free*** condition that to test whether a function is asymptotically convex; by checking the ***sign*** of $\int_{\mathbb{R}^n} f(\boldsymbol{x})\, d\boldsymbol{x}$. In addition, we prove that the minimizer for these asymptotically convex functions has a very simple closed form; it is the ***center of mass*** of the original function. Similar results about center of mass

and the sign of the integral are obtained by [25], but under the more restricted setting of compactly supported functions.

The other side of our contribution is the application of Gaussian smoothing and continuation to some vision problems. Specifically, we study the problem of geometric alignment of 2D images as well as 3D point clouds. The surface of the alignment objective for these problems typically has a lot local minima. Therefore, smoothing the optimization landscape seems a reasonable approach for these tasks. In order to smooth the alignment objective function, one would need to convolve it with a Gaussian kernel in the transformation space. Numerical computation of such integral can be expensive.

We will show that, for smoothing the objective of common motion models, such as affine and homography, there exists a corresponding ***integral operator on the image space*** . We refer to the kernels of such integral operators as ***transformation kernels*** . Thus, instead of convolving the objective function with a Gaussian kernel in transformation space, we can equivalently compute an integral transform in the image space. The former is very expensive to compute due to the effect of ***curse of dimensionality*** on numerical integration. For example, in case of 2D homography alignment, the transformation space is 8-dimensional, but the image space is 2-dimensional. As we show, all of these kernels are spatially varying as long as the transformation is not a pure translation, and vary from those heuristically suggested by [26] or [27].

Except for homography, the other kernels we derived can be expressed generically for signals with any dimensionality. This allows us to perform, for example affine alignment, using the same form of kernel on 2D images as well as 3D point clouds. We will show in our experimental results that, utilizing transformation kernels for aligning 2D and 3D data outperforms the traditional methods of plain gradient descent without any smoothing or Gaussian smoothing of the signals (instead of the objective function).

CHAPTER 2

ASYMPTOTIC ANALYSIS

Despite the long standing popularity of smoothing methods for optimization, there has been little theoretical work about it. This chapter rigorously studies some of the fundamental properties of the smoothing technique. In particular, we present a formal definition for the functions that can eventually become convex by smoothing. We refer to such functions as ***asymptotically convex functions*** . We show that, under mild conditions, asymptotically convex functions are a nontrivial superset of convex functions and inherit some of the nice properties of convex functions. For example, they obey some form of gradient inequality and positive-definite Hessian. We present an extremely simple and ***derivative-free*** condition to test whether a function is asymptotically convex; by checking the ***sign*** of $\int_{\mathbb{R}^n} f(\boldsymbol{x})\, d\boldsymbol{x}$. In addition, we prove that the minimizer for an asymptotically convex function has a very simple closed form; it is the ***center of mass*** of the original function. Similar results about center of mass and the sign of the integral are obtained by [25], but under the more restricted setting of compactly supported functions.

Admittedly, the asymptotic convexity, although a crucial necessary condition, is not sufficient to guarantee finding the global optimum through following the path of minimizers of the smoothed functions. The latter requires additional conditions on the objective function to ensure no singularity along the path being traced. In addition, it must guarantee that the path originated from the asymptotic minimizer eventually lands at the optimal (or ϵ-optimal) minimizer of the actual function.

2.1 Assumptions and Notations

Throughout this chapter, by smoothing the function f we mean convolving it with the isotropic Gaussian kernel $k(\boldsymbol{x};\sigma^2)$, with $\sigma > 0$, as defined below,

$$k(\boldsymbol{x};\sigma^2) \triangleq \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{\|\boldsymbol{x}\|^2}{2\sigma^2}} .$$

We also assume that the domain of f is the entire $\mathbb{R}^n$. The smoothed version of $f(\boldsymbol{x})$ is denoted by $g(\boldsymbol{x};\sigma)$,

$$g(\boldsymbol{x};\sigma) \triangleq f(\boldsymbol{x}) \star k(\boldsymbol{x};\sigma^2) = \int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\, d\boldsymbol{t}.$$

Note that such g is the solution to the ***heat equation*** [28] in the domain of $\mathbb{R}^n$ and with initial condition $g(\boldsymbol{x};0) = f(\boldsymbol{x})$.

Throughout the chapter, we always assume $f(\boldsymbol{x})$ has ***sub-exponential growth*** [1], i.e. it satisfies $\exists \rho \geq 0\,,\, \forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\boldsymbol{0},\rho)\;;\; |f(\boldsymbol{x})| < e^{\|\boldsymbol{x}\|}$. This is only to keep $g(\boldsymbol{x};\sigma)$, which is obtained by the Gaussian convolution, well-defined.

2.2 Definitions

Definition A real-valued continuous function $f(\boldsymbol{x})$ is called ***asymptotically convex*** if the following statement holds:

$$\begin{aligned}
&\forall M > 0\,,\, \exists\, \sigma^*_M\,,\, \forall \boldsymbol{x}_1 \in \mathcal{B}(\boldsymbol{0},M)\,,\, \forall \boldsymbol{x}_2 \in \mathcal{B}(\boldsymbol{0},M) \\
&\forall \lambda \in [0,1]\,,\, \forall \sigma \geq \sigma^*_M\;: \\
&g\big(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2;\sigma\big) \leq \lambda g(\boldsymbol{x}_1;\sigma) + (1-\lambda) g(\boldsymbol{x}_2;\sigma)\,. \qquad (2.1)
\end{aligned}$$

Definition An ***asymptotically strict convex*** function f is defined similar to the asymptotic convex function but with strict inequality in (2.1).

Example Consider the objective function of form $f(x) = e^{-\frac{x^2}{2\epsilon^2}} - e^{-\frac{x^2\epsilon^2}{2}}$ for $\epsilon > 0$. For small ϵ, this function looks like ℓ_0 norm, often seen in the literature of ***feature selection*** and ***sparse representation*** . The function f

[1] Note that the exponent is nonnegative in our definition of sub-exponential growth.

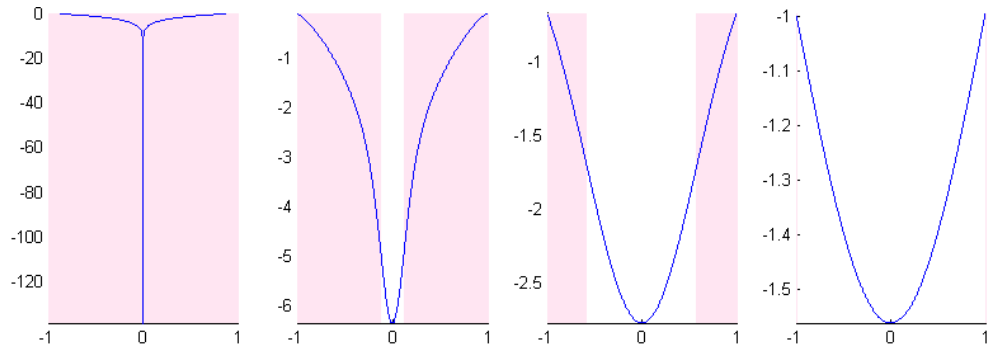


Figure 2.1: **Left to Right**: The function $g(x;\sigma) = [f \star k(\,.\,;\sigma^2)](x)$, where $f(x) = e^{-\frac{x^2}{2\epsilon^2}} - e^{-\frac{x^2\epsilon^2}{2}}$, with increasing values of σ. Nonconvex regions are colored by pink.

provides a much better surrogate for ℓ_0 norm compared to the ℓ_1 norm. Interestingly, while $f(x)$, except at its tip, is ***concave everywhere*** , it is asymptotically convex (Figure 2.1)!

2.3 Basic Properties

Asymptotically convex functions inherit the well-known gradient and Hessian properties of standard convex function. This will be shown in the following two propositions. The proofs are similar to their standard counterparts. However, here we present the proofs for completeness.

Interestingly, unlike non-asymptotic counterparts of gradient and Hessian inequalities, the asymptotic ones do "not" require once and twice differentiability conditions. That is because any function (with sub-exponential growth) convolved with a Gaussian kernel becomes C^∞, thus infinitely differentiable.

Proposition 1 (Gradient Inequality) *A function $f(\boldsymbol{x})$ is* ***asymptotically convex*** *"if and only if" it obeys the following gradient inequality:*

$$\begin{aligned}
&\forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x}_1 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \boldsymbol{x}_2 \in \mathcal{B}(\mathbf{0}, M) \\
&\forall \lambda \in [0,1]\,,\, \forall \sigma \geq \sigma_M^* \;: \\
&g(\boldsymbol{x}_2;\sigma) - g(\boldsymbol{x}_1;\sigma) \leq (\boldsymbol{x}_2 - \boldsymbol{x}_1)^T \nabla g(\boldsymbol{x}_2;\sigma)\,. \qquad (2.2)
\end{aligned}$$

Proof Sketch .

1. Supposing f is asymptotically convex, we prove the gradient inequality by writing the definition of asymptotic convexity (2.1) and setting $\lambda \to 0$.

2. Assuming the gradient inequality holds, we prove $g(\boldsymbol{x}; \sigma)$ is asymptotically convex by applying gradient inequality to the pair of points $\boldsymbol{x}_1$ and $\boldsymbol{x}_3$, as well as $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$, where $\boldsymbol{x}_3 = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2$. Taking convex combination of the two inequalities finishes the proof.

□

Proposition 2 (Hessian Condition for Asymptotic Convexity) *The function $f(\boldsymbol{x})$ is **asymptotically convex** "if and only if" it obeys the following condition:*

$$\begin{aligned} &\forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x} \in \mathcal{B}(\boldsymbol{0}, M)\,,\, \forall \sigma \geq \sigma_M^*\, : \\ &\nabla^2 g(\boldsymbol{x}; \sigma) \succeq \boldsymbol{O}\,. \end{aligned} \tag{2.3}$$

*Also, the function f is **asymptotically strict convex** under similar condition except that $\nabla^2 g(\boldsymbol{x}; \sigma) \succ \boldsymbol{O}$ instead of $\nabla^2 g(\boldsymbol{x}; \sigma) \succeq \boldsymbol{O}$.*

Proof Sketch .

We present the proof sketch for asymptotic convex case, and asymptotic strict convex case can be proved in a similar way.

1. Supposing f is asymptotically convex, we prove (2.3). We write g as its second order Taylor expansion of g at $\boldsymbol{x}$ along some arbitrary direction $\boldsymbol{u}$ plus Taylor's remainder (higher order terms). Since f is asymptotically convex, it obeys the gradient inequality (2.2). Applying that to the Taylor's expansion and observing that higher order terms can be ignored when $\lambda \to 0$, the result follows.

2. Assuming that (2.3) holds, we prove that f is asymptotically convex. Choose any pair of points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in $\mathcal{B}(\boldsymbol{0}, M)$. The proof uses a third point $\boldsymbol{x}_3 = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2$ for some $\lambda \in [0, 1]$ and then applies mean value theorem to derive the gradient inequality for asymptotically convex functions. Thus f is asymptotically convex.

□

Proposition 3 *Any convex function $f : \mathbb{R}^n \to \mathbb{R}$ is asymptotically convex.*

Proof Sketch . The proof simply starts with the definition of a convex function and then exploits the non-negativity of the Gaussian kernel. Integrating both sides of the inequality proves the proposition.

□

2.4 Derivative-Free Results for Decaying Functions

2.4.1 Asymptotic Convexity for Functions with Rapid Decay

Main Result (Corollary 9) *Consider a continuous $f : \mathbb{R}^n \to \mathbb{R}$. Suppose there exists an origin-centered ball, out of which f decays like $\|\boldsymbol{x}\|^{-n-3}$ or faster. If $\int_{\mathbb{R}}^{n} f(\boldsymbol{x})\, d\boldsymbol{x} < 0$, then f is asymptotically strict convex.*

Example *Show that $f(x) = -(\frac{\cos(x)}{1+x^4})^3$ is asymptotically convex.*

We first show that $f(x)$ decays fast enough. Choose $M^* = 0$ and $c = 1$, then it follows that $\forall x\,;\, |f(x)| \leq x^{-4}$. This can be checked by observing that $|f(x)| \leq |\cos(x)|(1+x^4)^{-3} \leq (1+x^4)^{-3} < x^{-4}$. Thus, it just remains to show $\int_{\mathbb{R}} f(x)\, d\boldsymbol{x} < 0$. Computing the closed form of $\int_{\mathbb{R}} f(x)\, dx$ is difficult. However, using the fact that $-\cos(x) \leq \frac{x^2}{2} - 1$, we construct an upper bound for $f(x)$ as $(\frac{\frac{x^2}{2}-1}{1+x^4})^3$. We have $\int_{\mathbb{R}} (\frac{\frac{x^2}{2}-1}{1+x^4})^3\, dx = -\frac{123\pi}{256\sqrt{2}} < 0$. Hence $f(x)$ is asymptotically convex.

The proof for the main result is now presented in a modular fashion through several pieces.

Proposition 4 *For any two real vectors $\boldsymbol{x}$ and $\boldsymbol{t}$, and any $\sigma > 0$, the following inequality holds:*

$$0 \leq 1 + \Big(\frac{(x_k - t_k)^2}{\sigma^2} - 1\Big) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \leq \frac{3}{2}\frac{(x_k - t_k)^2}{\sigma^2}.$$

Proof Sketch .

The key is the inequality $\frac{3}{2}z^2 - 1 \geq \big(z^2 - 1\big)e^{-\frac{z^2}{2}} \geq -1$ for any $z \in \mathbb{R}$. Choosing $z = (x_k - t_k)/\sigma$ and some manipulation proves the result.

□

Lemma 5 *Consider $f : \mathbb{R}^n \to \mathbb{R}$. Suppose there exists for f some $M^* \geq 0$, $c \geq 0$ and integer $a \geq n+3$ with the following property:*

$$\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)\,;\, |f(\boldsymbol{x})| \leq c\|\boldsymbol{x}\|^{-a}\ .$$

Then, the following inequality holds for any $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$:

$$\begin{aligned} & \Big| \int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t} \ + \ \sigma^2 (2\pi\sigma^2)^{\frac{n}{2}} \frac{\partial}{\partial x_k^2} g(\boldsymbol{x};\sigma) \Big| \\ \leq \quad & \frac{3}{2\sigma^2}(M+M^*)^2 \|f\| \ \Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \Big)^{\frac{1}{2}} \\ & + \frac{3cn\pi^{\frac{n}{2}} M^{*(n-a)}}{2\sigma^2 \Gamma(\frac{n}{2}+1)} \Big(\frac{M^2}{a-n} + 2\frac{MM^*}{a-n-1} + \frac{M^{*2}}{a-n-2} \Big). \end{aligned}$$

Proof Sketch .

We split the domain of integration within $\int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t}$ into interior and exterior of the ball $\mathcal{B}(\mathbf{0}, M^*)$ and upper bound each of them separately. The integral over the interior can be bounded using Cauchy-Schwartz inequality (to split f from k) and then applying Proposition 4 to the Gaussian kernel.

The exterior integral uses Proposition 4 and then applies lemma's assumption about the decay rate of f. This gives a ***radially symmetric*** integrand that is easy to compute. The result follows by putting together bounds on the interior and exterior integrals.

□

Proposition 6 *For any two real vectors $\boldsymbol{x}$ and $\boldsymbol{t}$, and any $\sigma > 0$, the following inequality holds:*

$$0 \leq \frac{|x_j - t_j|\ |x_k - t_k|}{\sigma^2} e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \leq \frac{|x_j - t_j|\ |x_k - t_k|}{\sigma^2}.$$

Proof Sketch .

We use the inequality $|z| \geq |z| e^{-\frac{z^2}{2}} \geq 0$ that holds for any $z \in \mathbb{R}$. In particular, choosing $z = (x_j - t_j)/\sigma$ and then again $z = (x_k - t_k)/\sigma$, plus some manipulation proves the result.

□

Lemma 7 *Consider $f : \mathbb{R}^n \to \mathbb{R}$. Suppose there exists for f some $M^* \geq 0$, $c \geq 0$ and integer $a \geq n+3$ with the following property:*

$$\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)\,;\, |f(\boldsymbol{x})| \leq c\|\boldsymbol{x}\|^{-a}.$$

Then, the following inequality holds for any $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$:

$$\begin{aligned} & \Big| \sigma^2(2\pi\sigma^2)^{\frac{n}{2}} \frac{\partial^2}{\partial x_j x_k} g(\boldsymbol{x};\sigma) \Big| \\ \leq\ & \frac{(M+M^*)^2}{\sigma^2}\|f\| \Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \Big)^{\frac{1}{2}} \\ & + \frac{cn\pi^{\frac{n}{2}} M^{*(n-a)}}{\sigma^2 \Gamma(\frac{n}{2}+1)} \Big(\frac{M^2}{a-n} + 2\frac{MM^*}{a-n-1} + \frac{M^{*2}}{a-n-2} \Big). \end{aligned}$$

Proof Sketch .

The proof strategy is very similar to that of Lemma 5.

□

Theorem 8 *Consider $f : \mathbb{R}^n \to \mathbb{R}$ such that $\|f\| \neq \infty$. Suppose there exists for f some $M^* \geq 0$, $c \geq 0$ and integer $a \geq n+3$ with the following property:*

$$\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)\,;\, |f(\boldsymbol{x})| \leq c\|\boldsymbol{x}\|^{-a}.$$

Then, for any $\epsilon > 0$ and any $M > 0$, if σ (as a function of ϵ and M) is chosen in the following sense,

$$\begin{aligned} \sigma^2 \ \geq\ & \frac{n}{\epsilon}(n+\frac{1}{2}) \Big(\|f\|(M^2+M^{*2}) \big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \big)^{\frac{1}{2}} \\ & + \frac{c\, n\, \pi^{\frac{n}{2}} M^{*(n-a)}}{\Gamma(\frac{n}{2}+1)} \cdot \\ & \big(\frac{M^2}{a-n} + 2\frac{MM^*}{a-n-1} + \frac{M^{*2}}{a-n-2} \big) \Big), \end{aligned}$$

then, it is guaranteed to have[2]*,*

$$\forall \boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M),$$
$$\Big\| \boldsymbol{I} \int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t} \;+\; \sigma^2(\sqrt{2\pi}\sigma)^n \nabla^2 g(\boldsymbol{x};\sigma) \Big\|_\infty \leq \epsilon\,.$$

Proof Sketch .

We first upper bound $\|\boldsymbol{I} \int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t} + \sigma^2(\sqrt{2\pi}\sigma)^n \nabla^2 g(\boldsymbol{x};\sigma)\|_\infty$ by applying Lemma 5 to the diagonals and Lemma 7 to the off-diagonals of the expression inside $\|\ .\ \|_\infty$. In order to prove the theorem, it is sufficient to set the derived upper bound less than ϵ.

□

Corollary 9 *Consider a continuous function $f : \mathbb{R}^n \to \mathbb{R}$. Suppose there exists for f some $M^* \geq 0$, $c \geq 0$ and integer $a \geq n+3$ with the following property,*

$$\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)\,;\, |f(\boldsymbol{x})| \leq c\|\boldsymbol{x}\|^{-a}\,. \tag{2.4}$$

If $\int_{\mathbb{R}^n} f(\boldsymbol{x})\, d\boldsymbol{x} < 0$ then f is asymptotically strict convex.
More precisely, it will hold that,

$$\forall M > 0\,,\, \exists\, \sigma^*_M\,,\, \forall \boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \sigma \geq \sigma^*_M\ :$$
$$\nabla^2 g(\boldsymbol{x};\sigma) \succ \boldsymbol{O}\,, \tag{2.5}$$

where σ^*_M is defined as below,

$$\sigma^{*\,2}_M \;\triangleq\; \frac{n^2(n+\frac{1}{2})}{-\int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t}} \Bigg(\|f\|(M^2 + M^{*2}) \Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \Big)^{\frac{1}{2}} \tag{2.6}$$
$$+ \frac{c\, n\, \pi^{\frac{n}{2}} M^{*(n-a)}}{\Gamma(\frac{n}{2}+1)} \Big(\frac{M^2}{a-n} + 2\frac{MM^*}{a-n-1} + \frac{M^{*2}}{a-n-2} \Big) \Bigg)\,. \tag{2.7}$$

Proof Sketch .

Considering Proposition 2, any function f is asymptotically strict convex if and only if for any $M > 0$ there exists a σ^*_M, such that for any $\sigma \geq$

[2]The notation $\|\boldsymbol{A}\|_\infty$ is for the max-norm of the matrix $\boldsymbol{A}$ and is defined as $\|\boldsymbol{A}\|_\infty = \max_{i,j} |a_{ij}|$.

σ_M^*, $\nabla^2 g(\boldsymbol{x};\sigma)$ or equivalently $\sigma^2(\sqrt{2\pi}\sigma)^n\nabla^2 g(\boldsymbol{x};\sigma)$ is positive definite within $\mathcal{B}(\mathbf{0}, M)$.

The tail decay assumption allows us to apply Theorem 8. Note that here $\|f\| < \infty$ due to its decay rate and continuity of f. Thus, $\|\boldsymbol{I}\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} + \sigma^2(\sqrt{2\pi}\sigma)^n\nabla^2 g(\boldsymbol{x};\sigma)\|_\infty$ can be made arbitrarily small. Consequently, if $\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} < 0$, then $\sigma^2(\sqrt{2\pi}\sigma)^n\nabla^2 g(\boldsymbol{x};\sigma)$ can be made arbitrarily close to a matrix with zero off-diagonals and strictly positive diagonals and thus f is asymptotically convex.

□

2.4.2 Asymptotic Minimizer for Functions with Rapidly Enough Decay Rate

Main Result (Corollary 13) *Consider $f : \mathbb{R}^n \to \mathbb{R}$. Suppose there is an origin-centered ball out of which f decays like $\|\boldsymbol{x}\|^{-n-4}$ or faster. Then for any origin-centered ball, there always exists some σ that can make $\frac{\int_{\mathbb{R}^n} \boldsymbol{t}\, f(\boldsymbol{t})\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}}$ arbitrary close to the stationary point of $g(\boldsymbol{x};\sigma)$.*

Example *Show that $f(x) = e^{-\frac{(x-1)^2}{0.1}} - e^{-x^2}$ is asymptotically convex and find its asymptotic minimizer.*

It is easy to check that $f(x)$ satisfies the decay condition because $\forall x|f(x)| \leq 2|x|^{-5}$. On the other hand, $\int_{\mathbb{R}} f(x)\,dx = -1.21195$ (up to 6 decimal digits). Therefore, $f(x)$ is asymptotically convex. In addition $x^* = \frac{\int_{\mathbb{R}} x f(x) dx}{\int_{\mathbb{R}} f(x)\,dx} = -0.46247$ (up to 6 decimal digits). Figure 2.2 shows how f looks more convex and its minimizer approaches -0.46247 as σ increases.

In the following, we present a modular proof of the main result.

Lemma 10 (Zeroth Moment Convergence) *Consider $f : \mathbb{R}^n \to \mathbb{R}$. Suppose there exists for f some $M^* \geq 0$, $c \geq 0$ and integer $a \geq n+3$ with the following property:*

$$\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)\,;\, |f(\boldsymbol{x})| \leq c\|\boldsymbol{x}\|^{-a}.$$

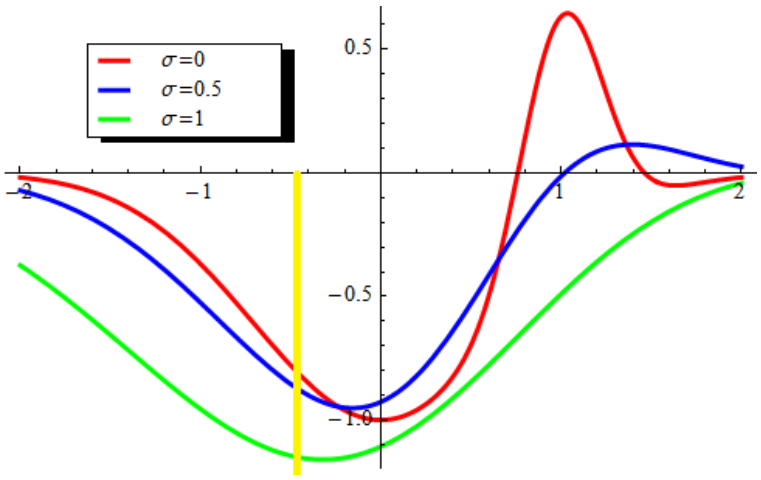


Figure 2.2: $g(x;\sigma) = f(x) \star k(x;\sigma)$, $f(x) = e^{-\frac{(x-1)^2}{0.1}} - e^{-x^2}$ for different choices of σ. Notice that the minimizer becomes unique for large σ and approaches the yellow line $x = -0.46247$.

Then for any $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$ *the following inequality holds:*

$$\begin{aligned} & \Big| \int_{\mathbb{R}^n} f(\boldsymbol{t}) \Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big) \, d\boldsymbol{t} \Big| \\ \leq \; & \frac{(M+M^*)^2}{2\sigma^2} \|f\| \, \big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \big)^{\frac{1}{2}} \\ & + \frac{cn\pi^{\frac{n}{2}} M^{*n-a}}{2\sigma^2 \Gamma(\frac{n}{2}+1)} \big(\frac{M^2}{a-n} + 2\frac{MM^*}{a-n-1} + \frac{M^{*2}}{a-n-2} \big). \end{aligned}$$

Proof Sketch .

The proof is similar to that of Lemma 5, except that here we use an additional fact $\forall y \in \mathbb{R} \; ; \; 1 - e^{-y^2} \leq y^2$ when bounding the interior integral. □

Lemma 11 (First Moment Convergence) *Consider* $f : \mathbb{R}^n \to \mathbb{R}$. *Suppose there exists for* f *some* $M^* \geq 0$, $c \geq 0$ *and integer* $a \geq n+4$ *with the following property:*

$$\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*) \; ; \; |f(\boldsymbol{x})| \leq c\|\boldsymbol{x}\|^{-a}.$$

Define $h_i(\boldsymbol{t}) \triangleq t_i f(\boldsymbol{t})$. *Then for any* $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$, *and any* $i = 1, 2, \ldots, n$,

the following inequality holds:

$$\begin{aligned} & \Big| \int_{\mathbb{R}^n} h_i(\boldsymbol{t}) \Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big) \, d\boldsymbol{t} \Big| \\ \leq \; & \frac{(M+M^*)^2}{2\sigma^2} \|h_i\| \Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \Big)^{\frac{1}{2}} \\ & \cdot \frac{c\, n \pi^{\frac{n}{2}} M^{*n+1-a}}{2\sigma^2 \, \Gamma(\frac{n}{2}+1)} \\ & \cdot \Big(\frac{M^2}{a-n-1} + 2\frac{MM^*}{a-n-2} + \frac{M^{*2}}{a-n-3} \Big). \end{aligned}$$

Proof Sketch .

The proof strategy is very similar to that of 5, except that here we use an additional fact $\forall y \in \mathbb{R} \; ; \; 1 - e^{-y^2} \leq y^2$ when bounding the interior integral. □

Theorem 12 *Consider a continuous function $f : \mathbb{R}^n \to \mathbb{R}$. Suppose there exists for f some $M^* \geq 0$, $c \geq 0$ and integer $a \geq n+4$ with the following property:*

$$\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*) \; ; \; |f(\boldsymbol{x})| \leq c \|\boldsymbol{x}\|^{-a}.$$

Then, for any $\epsilon > 0$ and any $i = 1, 2, \ldots, n$, there always exists some $\sigma^* > 0$ such that for any $\sigma \geq \sigma^*$, and for any $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$ the following inequality holds:

$$\Big| \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i k(\boldsymbol{x}-\boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x}-\boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}} \Big| \leq \epsilon.$$

Proof Sketch .

The theorem seeks to bound $\Big| \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i k(\boldsymbol{x}-\boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x}-\boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}} \Big|$ or equivalently bound $\Big| \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \, d\boldsymbol{t}} \Big|$.

We first must ensure that the denominators are non-zero. We already know from theorem's assumptions that $\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t} \neq 0$. Thus, we just need to ensure $\int_{\mathbb{R}^n} f(\boldsymbol{t}) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \, d\boldsymbol{t} \neq 0$. We use the fact that $\forall (a,b) \in \mathbb{R} - \{0\} \times \mathbb{R} \; ; \; |a| \geq \mathrm{UB}(2|a-b|) \Rightarrow 2|b| \geq |a|$ where UB(.) is any function satisfying $\mathrm{UB}(x) \geq x$.

Therefore, in order to have $| \int_{\mathbb{R}^n} f(\boldsymbol{t}) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \, d\boldsymbol{t} | \geq \frac{1}{2} | \int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t} | \neq 0$, it is sufficient to satisfy $\frac{1}{2} | \int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t} | \geq \mathrm{UB}(| \int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t} - \int_{\mathbb{R}^n} f(\boldsymbol{t}) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \, d\boldsymbol{t}|)$. The $\mathrm{UB}(\cdot)$ here can be obtained using Lemma 10. It is easy to show that

there always exists some σ that satisfies the above sufficient condition, simply by moving σ to one side of the inequality and showing the other side is bounded (σ^* is the value of σ when inequality is replaced by equality). The boundedness holds because $\int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t} \neq 0$ by theorem's assumptions, and that $\|f\| < \infty$ due to its decay rate property.

After making sure that $\left| \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i\, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\, d\boldsymbol{t}} \right|$ is well-defined, we proceed to establish an upper bound for it. We use the fact that $\frac{a}{b} - \frac{c}{d} = a\frac{d-b}{bd} + \frac{a-c}{d}$, which implies that $|\frac{a}{b} - \frac{c}{d}| \leq \frac{|a|}{|b|}\frac{\mathrm{UB}(|d-b|)}{\mathrm{LB}(|d|)} + \frac{\mathrm{UB}(|a-c|)}{\mathrm{LB}(|d|)}$. Applying this fact to the expression we want to bound, and then moving σ to one side, we observe that such σ always exists, because the other side is always bounded (σ^* is obtained at the equality). More precisely, $\int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t} \neq 0$ due to theorem's assumptions, and $\|f\| < \infty$ and $\|h_i\| < \infty$ due to decay rate property of f.

Remember, however, we earlier had an additional constraint on σ^* to keep the denominator non-zero. In order for σ^* to jointly satisfy both conditions, we take the maximum of the two, which is still a bounded number and thus always exists.

□

Corollary 13 *Consider $f : \mathbb{R}^n \to \mathbb{R}$. Suppose there exists for f some $M^* \geq 0$, $c \geq 0$ and integer $a \geq n+4$ with the following property:*

$$\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)\,;\, |f(\boldsymbol{x})| \leq c\|\boldsymbol{x}\|^{-a}.$$

Let $\boldsymbol{x}^*_\sigma$ denote a stationary point of $g(\boldsymbol{x};\sigma)$, that is $\nabla g(\boldsymbol{x}^*_\sigma;\sigma) = \mathbf{0}$. Then, for any $\epsilon > 0$ and any $M \geq 0$, there always exists some (large enough) $\sigma > 0$ (which depends on ϵ and M) that can make $\|\frac{\int_{\mathbb{R}^n} \boldsymbol{t}\, f(\boldsymbol{t})\, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t}} - \boldsymbol{x}^*_\sigma\|_\infty$ arbitrarily small.

Proof Sketch .

The assumption on f allows application of Theorem 12. Consequently, the theorem holds for all $i = 1, 2, \cdots n$ "simultaneously", when stated as the following. For $i = 1, 2, \ldots, n$, and any $\epsilon_i > 0$ there "always exists" some $\sigma^* > 0$ such that for any $\sigma \geq \sigma^*$, and for any $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$ the following

inequality holds:

$$\begin{aligned}
&\forall i \in \{1, 2, \ldots, n\} \\
&\Big| \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}} \Big| \leq \epsilon \quad \Rightarrow \\
&\sum_{i=1}^{n} \Big| \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) d\boldsymbol{t}} \Big| \leq n\epsilon_i \, .
\end{aligned} \tag{2.8}$$

On the other hand, we know that $\|\boldsymbol{x}\|_\infty \leq \|\boldsymbol{x}\|_1$, which combined with (2.8) gives the following:

$$\Big\| \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}} \Big\|_\infty \leq n\epsilon \, . \tag{2.9}$$

We stress that for any "arbitrarily small" $\epsilon > 0$, there always exists some corresponding σ^* that satisfies the bound in (2.9) for any $\sigma \geq \sigma^*$.

It just remains to show how (2.9) is related to a stationary point of $g(\boldsymbol{x}; \sigma)$. We proceed by writing down the definition of a stationary point $\boldsymbol{x}^*_\sigma$ as $\nabla g(\boldsymbol{x}^*_\sigma; \sigma) = \int_{\mathbb{R}^n} f(\boldsymbol{t}) \frac{(\boldsymbol{x}^*_\sigma - \boldsymbol{t})}{\sigma^2} k(\boldsymbol{x}^*_\sigma - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}$. Zero crossing this expression leads to $\boldsymbol{x}^*_\sigma = \frac{\int_{\mathbb{R}^n} \boldsymbol{t} \, f(\boldsymbol{t}) k(\boldsymbol{x}^*_\sigma - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x}^*_\sigma - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}$. Plugging the latter into (2.9) proves the corollary.

□

2.5 The Function Space over $\mathbb{R}^n$

Consider the space of functions $\{f : \mathbb{R}^n \to \mathbb{R}\}$. Based on the materials presented in this chapter, it is easy to see how different subsets of such function space are related to each other (Figure 2.3). Remember, we defined a function to have sub-exponential growth if it satisfies $\exists \rho \geq 0 \, , \, \forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\boldsymbol{0}, \rho) \; ; \; |f(\boldsymbol{x})| < e^{\|\boldsymbol{x}\|}$ (note that the sign of the exponential is positive).

- From Proposition 3, we know that any convex function with sub-exponential growth is also asymptotically convex. For example, the convex function $f(x) = x^2$, when convolved with the Gaussian gives $g(x; \sigma) = x^2 + \sigma^2$. The latter is convex in x for any σ, including $\sigma \to \infty$, thus it is asymptotically convex. Hence, as long as we limit our focus

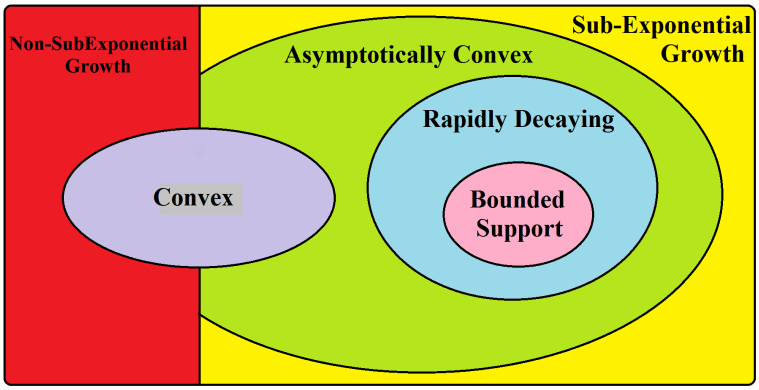


Figure 2.3: The taxonomy of the space of functions $\{f : \mathbb{R}^n \to \mathbb{R}\}$.

to functions with sub-exponential growth, asymptotically convex functions form a ***superset*** of standard convex functions. Of course, the reverse is not necessarily true; there exist convex functions that do not satisfy the growth condition, e.g. $f(x) = \exp(\exp(x))$.

- The class of asymptotically convex functions is very rich. Specifically, a lot of optimization problems can be equivalently expressed as one that is asymptotically convex. To see that, consider any (possibly nonconvex) unconstrained optimization problem $\{\boldsymbol{x}^*\} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$. If for any such $\boldsymbol{x}^*$, we have $|\boldsymbol{x}^*| \nrightarrow \infty$, then we argue that another problem can be constructed which has same set of minimizers, but is also asymptotically convex.

  The assumption $|\boldsymbol{x}^*| \nrightarrow \infty$ means there exists a large enough ball of radius ρ that will contain the (set of) $\{\boldsymbol{x}^*\}$. Let γ be such that $\forall \boldsymbol{x} \in \mathcal{B}(\mathbf{0}, \rho)\,;\, \gamma \geq f(\boldsymbol{x})$. Then the function $\hat{f}(\boldsymbol{x})$ as defined in (2.10) has the same minimizer as that of $f(\boldsymbol{x})$, i.e. $\{\boldsymbol{x}^*\} = \{\hat{\boldsymbol{x}}^*\}$, simply because they are different up to a constant. However, by Corollary 9, it is easy to check that $\hat{f}(\boldsymbol{x})$ is asymptotically convex.

$$\hat{f}(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}) - \gamma & \boldsymbol{x} \in \mathcal{B}(\mathbf{0}, \rho) \\ 0 & \text{otherwise} \end{cases} . \tag{2.10}$$

2.6 Proofs

Proposition 1 (Gradient Inequality for Asymptotic Convexity) *A function $f : \mathbb{R}^n \to \mathbb{R}$ is asymptotically convex "if and only if" it obeys the following gradient inequality.*

$$\forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x}_1 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \boldsymbol{x}_2 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \lambda \in [0,1]\,,\, \forall \sigma \geq \sigma_M^* \,:$$
$$g(\boldsymbol{x}_2;\sigma) - g(\boldsymbol{x}_1;\sigma) \leq (\boldsymbol{x}_2 - \boldsymbol{x}_1)^T \nabla g(\boldsymbol{x}_2;\sigma)\,. \tag{2.11}$$

Proof 1. We suppose f is asymptotically convex and prove that the gradient inequality holds. We start by writing the definition of asymptotic convexity (2.1) as below:

$$\begin{aligned}&\forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x}_1 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \boldsymbol{x}_2 \in \mathcal{B}(\mathbf{0}, M)\\ &\forall \lambda \in [0,1]\,,\, \forall \sigma \geq \sigma_M^* \,:\\ &g\big(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2;\sigma\big) \leq \lambda g(\boldsymbol{x}_1;\sigma) + (1-\lambda) g(\boldsymbol{x}_2;\sigma)\,.\end{aligned} \tag{2.12}$$

This implies that for $\lambda \in (0,1]$ we have,

$$\frac{g\big(\lambda(\boldsymbol{x}_1 - \boldsymbol{x}_2) + \boldsymbol{x}_2;\sigma\big) \,-\, g(\boldsymbol{x}_2;\sigma)}{\lambda} \,\leq\, g(\boldsymbol{x}_1;\sigma) \,-\, g(\boldsymbol{x}_2;\sigma)\,. \tag{2.13}$$

In particular, letting $\lambda \to 0$, using the definition of directional derivative, we derive the following,

$$(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T \nabla g(\boldsymbol{x}_2;\sigma) \,\leq\, g(\boldsymbol{x}_1;\sigma) \,-\, g(\boldsymbol{x}_2;\sigma) \tag{2.14}$$
$$\equiv\quad g(\boldsymbol{x}_2;\sigma) \,-\, g(\boldsymbol{x}_1;\sigma) \,\leq\, (\boldsymbol{x}_2 - \boldsymbol{x}_1)^T \nabla g(\boldsymbol{x}_2;\sigma)\,. \tag{2.15}$$

2. Now suppose that the gradient inequality holds. We prove that this implies $g(\boldsymbol{x};\sigma)$ is asymptotically convex. By the gradient inequality we have the following for any pair of points $\boldsymbol{x}_1$ and $\boldsymbol{x}_3$:

$$\forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x}_1 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \boldsymbol{x}_3 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \lambda \in [0,1]\,,\, \forall \sigma \geq \sigma_M^* \;:$$
$$g(\boldsymbol{x}_3;\sigma) - g(\boldsymbol{x}_1;\sigma) \leq (\boldsymbol{x}_3 - \boldsymbol{x}_1)^T \nabla g(\boldsymbol{x}_3;\sigma)\,. \tag{2.16}$$

Also, we can have the following for any pair of points $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$.

$$\forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x}_2 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \boldsymbol{x}_3 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \lambda \in [0,1]\,,\, \forall \sigma \geq \sigma_M^* \;:$$
$$g(\boldsymbol{x}_3;\sigma) - g(\boldsymbol{x}_2;\sigma) \leq (\boldsymbol{x}_3 - \boldsymbol{x}_2)^T \nabla g(\boldsymbol{x}_3;\sigma)\,. \tag{2.17}$$

In particular, we can choose $\boldsymbol{x}_3 \in \mathcal{B}(\mathbf{0}, M)$ such that it satisfies the following relationship.

$$\boldsymbol{x}_3 = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \tag{2.18}$$

Note that there exists such $\boldsymbol{x}_3 \in \mathcal{B}(\mathbf{0}, M)$ because $\mathcal{B}(\mathbf{0}, M)$ is a convex set and we know that any convex combination of two points within a convex set lies inside that set,

$$\begin{aligned}
&\forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x}_1 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \boldsymbol{x}_2 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \lambda \in [0,1]\,,\, \forall \sigma \geq \sigma_M^* \\
&\exists\, \boldsymbol{x}_3 \in \mathcal{B}(\mathbf{0}, M) \;: \\
&\boldsymbol{x}_3 = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \\
&g(\boldsymbol{x}_3;\sigma) - g(\boldsymbol{x}_1;\sigma) \leq (\boldsymbol{x}_3 - \boldsymbol{x}_1)^T \nabla g(\boldsymbol{x}_3;\sigma) \\
&g(\boldsymbol{x}_3;\sigma) - g(\boldsymbol{x}_2;\sigma) \leq (\boldsymbol{x}_3 - \boldsymbol{x}_2)^T \nabla g(\boldsymbol{x}_3;\sigma)\;.
\end{aligned} \tag{2.19}$$

In particular, taking the convex combination of the above inequalities implies below,

$$
\begin{aligned}
&\forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x}_1 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \boldsymbol{x}_2 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \lambda \in [0,1]\,,\, \forall \sigma \geq \sigma_M^* \\
&\exists\, \boldsymbol{x}_3 \in \mathcal{B}(\mathbf{0}, M)\ : \\
&\boldsymbol{x}_3 = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \\
&\quad \lambda g(\boldsymbol{x}_3;\sigma) - \lambda g(\boldsymbol{x}_1;\sigma) + (1-\lambda) g(\boldsymbol{x}_3;\sigma) - (1-\lambda) g(\boldsymbol{x}_2;\sigma) \\
&\leq \lambda (\boldsymbol{x}_3 - \boldsymbol{x}_1)^T \nabla g(\boldsymbol{x}_3;\sigma) + (1-\lambda)(\boldsymbol{x}_3 - \boldsymbol{x}_2)^T \nabla g(\boldsymbol{x}_3;\sigma)\ . \qquad (2.20)
\end{aligned}
$$

We now plug in $\boldsymbol{x}_3 = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2$ into RHS of the inequality (2.20) and manipulate it as below.

$$
\begin{aligned}
& \lambda (\boldsymbol{x}_3 - \boldsymbol{x}_1)^T \nabla g(\boldsymbol{x}_3;\sigma) + (1-\lambda)(\boldsymbol{x}_3 - \boldsymbol{x}_2)^T \nabla g(\boldsymbol{x}_3;\sigma) \\
= \;& \Big(\lambda(\boldsymbol{x}_3 - \boldsymbol{x}_1) + (1-\lambda)(\boldsymbol{x}_3 - \boldsymbol{x}_2)\Big)^T \nabla g(\boldsymbol{x}_3;\sigma) \\
= \;& \Big(\lambda(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 - \boldsymbol{x}_1) + (1-\lambda)(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 - \boldsymbol{x}_2)\Big)^T \nabla g(\boldsymbol{x}_3;\sigma) \\
= \;& \mathbf{0}^T \nabla g(\boldsymbol{x}_3;\sigma)\,. \qquad (2.21)
\end{aligned}
$$

Therefore, (2.20) can be restated as below:

$$
\begin{aligned}
&\forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x}_1 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \boldsymbol{x}_2 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \lambda \in [0,1]\,,\, \forall \sigma \geq \sigma_M^* \\
&\exists\, \boldsymbol{x}_3 \in \mathcal{B}(\mathbf{0}, M)\ : \\
&\boldsymbol{x}_3 = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \\
&g(\boldsymbol{x}_3;\sigma) - \lambda g(\boldsymbol{x}_1;\sigma) - (1-\lambda) g(\boldsymbol{x}_2;\sigma) \leq 0\ . \qquad (2.22)
\end{aligned}
$$

Finally, plugging the definition of $\boldsymbol{x}_3$ again into (2.22), we derive:

$$
\begin{aligned}
&\forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x}_1 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \boldsymbol{x}_2 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \lambda \in [0,1]\,,\, \forall \sigma \geq \sigma_M^* \\
&g\Big(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2;\sigma\Big) - \lambda g(\boldsymbol{x}_1;\sigma) - (1-\lambda) g(\boldsymbol{x}_2;\sigma) \leq 0\ . \qquad (2.23)
\end{aligned}
$$

□

Proposition 2 (Hessian Condition for Asymptotic Convexity) *The function $f(\boldsymbol{x})$ is asymptotically convex "if and only if" it obeys the following condition.*

$$\begin{aligned} &\forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \sigma \geq \sigma_M^* \;: \\ &\nabla^2 g(\boldsymbol{x}; \sigma) \succeq \boldsymbol{O}\,. \end{aligned} \tag{2.24}$$

Also, the function f is ***asymptotically strict convex*** *under similar conditions except that $\nabla^2 g(\boldsymbol{x}; \sigma) \succeq \boldsymbol{O}$ instead of $\nabla^2 g(\boldsymbol{x}; \sigma) \succ \boldsymbol{O}$.*

Proof We present the proof for asymptotic convex case, and the asymptotic "strict" convex case can be proved in a similar way.

1. We suppose f is asymptotically convex and then prove it implies the following condition of the Proposition,

$$\begin{aligned} &\forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \sigma \geq \sigma_M^* \;: \\ &\nabla^2 g(\boldsymbol{x}; \sigma) \succeq \boldsymbol{O}\,. \end{aligned} \tag{2.25}$$

Let $\boldsymbol{u} \in \mathbb{R}^n$, such that $\|\boldsymbol{u}\| = 1$, be any direction. Using second order Taylor's expansion of g around $\boldsymbol{x}$, we have the following:

$$g\big(\boldsymbol{x} + \lambda \boldsymbol{u}; \sigma\big) \;=\; g\big(\boldsymbol{x}; \sigma\big) \;+\; \lambda \boldsymbol{u}^T \nabla g\big(\boldsymbol{x}; \sigma\big) \;+\; \frac{1}{2}\lambda^2 \boldsymbol{u}^T \nabla^2 g\big(\boldsymbol{x}; \sigma\big)\boldsymbol{u} \;+\; r(\boldsymbol{x}, \lambda \boldsymbol{u})\,, \tag{2.26}$$

where $r(\boldsymbol{x}, \lambda \boldsymbol{u})$ is the remainder of the Taylor's expansion. Since f is asymptotically convex, it obeys the gradient inequality (2.2) (with $\boldsymbol{x}_1$ being $\boldsymbol{x} + \lambda \boldsymbol{u}$ and $\boldsymbol{x}_2$ being $\boldsymbol{x}$ here). Therefore, (2.26) can be rewritten as below.

$$\begin{aligned} &\forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \sigma \geq \sigma_M^* \;: \\ &\frac{1}{2}\lambda^2 \boldsymbol{u}^T \nabla^2 g\big(\boldsymbol{x}; \sigma\big)\boldsymbol{u} \;+\; r(\boldsymbol{x}, \lambda \boldsymbol{u}) \geq 0\,. \end{aligned} \tag{2.27}$$

In particular, letting $\lambda \to 0$, $|r(\boldsymbol{x}, \lambda \boldsymbol{u})| \ll |\frac{1}{2}\lambda^2 \boldsymbol{u}^T \nabla^2 g\big(\boldsymbol{x}; \sigma\big)\boldsymbol{u}|$, we derive $\boldsymbol{u}^T \nabla^2 g\big(\boldsymbol{x}; \sigma\big)\boldsymbol{u} \geq 0$, which is equivalent to $\nabla^2 g\big(\boldsymbol{x}; \sigma\big) \succeq \boldsymbol{O}$.

2. We assume that $\forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \sigma \geq \sigma_M^* \Rightarrow \nabla^2 g(\boldsymbol{x}; \sigma) \succeq \boldsymbol{O}$ holds and prove that f is asymptotically convex.

Choose any pair of points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in $\mathcal{B}(\mathbf{0}, M)$. Then there exists a point $\boldsymbol{x}_3 = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2$ for some $\lambda \in [0,1]$ such that the following identity holds (an extension of mean value theorem to the second derivative).

$$g(\boldsymbol{x}_1; \sigma) = g(\boldsymbol{x}_2; \sigma) + (\boldsymbol{x}_1 - \boldsymbol{x}_2)^T \nabla g(\boldsymbol{x}_2; \sigma) + (\boldsymbol{x}_1 - \boldsymbol{x}_2)^T \nabla^2 g(\boldsymbol{x}_3; \sigma)(\boldsymbol{x}_1 - \boldsymbol{x}_2)\,. \tag{2.28}$$

However, since $\nabla^2 g(\boldsymbol{x}; \sigma) \succeq \boldsymbol{O}$ for any $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$, it holds at $\boldsymbol{x} = \boldsymbol{x}_3$ as well. Note that $\boldsymbol{x}_3$ has to lie inside the convex set $\mathcal{B}(\mathbf{0}, M)$ because $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are in the latter ball, and $\boldsymbol{x}_3$ is merely a convex combination of these two points. Thus, we derive the following inequality,

$$\begin{aligned} & \forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x}_1 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \boldsymbol{x}_2 \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \lambda \in [0,1]\,,\, \forall \sigma \geq \sigma_M^* \,: \\ & g(\boldsymbol{x}_1; \sigma) - g(\boldsymbol{x}_2; \sigma) - (\boldsymbol{x}_1 - \boldsymbol{x}_2)^T \nabla g(\boldsymbol{x}_2; \sigma) \geq 0 \qquad (2.29) \\ \equiv \quad & g(\boldsymbol{x}_2; \sigma) - g(\boldsymbol{x}_1; \sigma) \leq (\boldsymbol{x}_2 - \boldsymbol{x}_1)^T \nabla g(\boldsymbol{x}_2; \sigma)\,. \qquad (2.30) \end{aligned}$$

The above is the gradient inequality (2.2) for asymptotically convex functions and thus f is asymptotically convex.

□

Proposition 3 *Any convex function with sub-exponential growth is asymptotically convex.*

Proof The proof simply starts with the definition of a convex function and then exploits the non-negativity of the Gaussian kernel. Sub-Exponential growth condition is merely used to keep Gaussian convolution bounded and hence well-defined.

$$
\begin{aligned}
& \forall \lambda \in [0,1]\ ,\ \forall (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{t}) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n\ : \\
& f\Big(\lambda(\boldsymbol{x}_1 - \boldsymbol{t}) + (1-\lambda)(\boldsymbol{x}_2 - \boldsymbol{t})\Big) \leq \lambda f(\boldsymbol{x}_1 - \boldsymbol{t}) + (1-\lambda) f(\boldsymbol{x}_2 - \boldsymbol{t}) \\
\Rightarrow \quad & \int_{\mathbb{R}^n} k(\boldsymbol{t};\sigma) f\Big(\lambda(\boldsymbol{x}_1 - \boldsymbol{t}) + (1-\lambda)(\boldsymbol{x}_2 - \boldsymbol{t})\Big) d\boldsymbol{t} \\
\leq \quad & \int_{\mathbb{R}^n} k(\boldsymbol{t};\sigma) \Big(\lambda f(\boldsymbol{x}_1 - \boldsymbol{t}) + (1-\lambda) f(\boldsymbol{x}_2 - \boldsymbol{t})\Big) d\boldsymbol{t} \\
\equiv \quad & g\Big(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2; \sigma\Big) \leq \lambda g(\boldsymbol{x}_1;\sigma) + (1-\lambda) g(\boldsymbol{x}_2;\sigma)\,. \qquad (2.31)
\end{aligned}
$$

This result is independent of M and thus holds for any arbitrary value of M for satisfying the definition (2.1).

□

2.6.1 Derivative Free Results on Asymptotic Convexity

Main Result (Corollary 9) *Consider a continuous $f : \mathbb{R}^n \to \mathbb{R}$. Suppose there exists an origin-centered ball, out of which f decays like $\|\boldsymbol{x}\|^{-n-3}$ or faster. If $\int_{\mathbb{R}}^n f(\boldsymbol{x})\, d\boldsymbol{x} < 0$, then f is asymptotically strict convex.*

In the sequel, we present the proof of the main result.

Proposition 4 (Bound on Hessian's Diagonal of a Gaussian) *For any real vectors $\boldsymbol{x}$ and $\boldsymbol{t}$ and any $\sigma^2 > 0$, the following inequality holds:*

$$
0 \leq 1 + \big(\frac{(x_k - t_k)^2}{\sigma^2} - 1\big) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \leq \frac{3}{2}\frac{(x_k - t_k)^2}{\sigma^2}\,. \qquad (2.32)
$$

Proof Consider the function $\big(z^2 - 1\big)e^{-\frac{z^2}{2}}$ defined for any $z \in \mathbb{R}$. We first obtain an over-estimator for this function that has simpler form. In fact, it turns out the Taylor expansion up to the second order term gives such an over estimator:

$$
\forall z \in \mathbb{R} \quad ; \quad \frac{3}{2}z^2 - 1 \geq \big(z^2 - 1\big)e^{-\frac{z^2}{2}} \geq -1\,. \qquad (2.33)
$$

In particular, by choosing $z = (x_k - t_k)/\sigma$, we can proceed as below.

$$
\begin{aligned}
& \frac{3}{2}\frac{(x_k-t_k)^2}{\sigma^2} - 1 \geq \big(\frac{(x_k-t_k)^2}{\sigma^2} - 1\big) e^{-\frac{(x_k-t_k)^2}{2\sigma^2}} \geq -1 \\
\Rightarrow \quad & \frac{3}{2}\frac{(x_k-t_k)^2}{\sigma^2} - 1 \geq \big(\frac{(x_k-t_k)^2}{\sigma^2} - 1\big) e^{-\frac{(x_k-t_k)^2}{2\sigma^2}} e^{-\frac{\sum_{i\neq k}(x_i-t_i)^2}{2\sigma^2}} \geq -1 \\
\equiv \quad & \frac{3}{2}\frac{(x_k-t_k)^2}{\sigma^2} - 1 \geq \big(\frac{(x_k-t_k)^2}{\sigma^2} - 1\big) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \geq -1 \\
\equiv \quad & \frac{3}{2}\frac{(x_k-t_k)^2}{\sigma^2} \geq \big(\frac{(x_k-t_k)^2}{\sigma^2} - 1\big) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} + 1 \geq 0\,. \qquad (2.34)
\end{aligned}
$$

□

Lemma 5 (Convergence of Hessian's Diagonal) *Consider* $f : \mathbb{R}^n \to \mathbb{R}$. *Suppose there exists for* f *some* $M^* \geq 0$, $c \geq 0$ *and integer* $a \geq n+3$ *with the following property:*

$$
\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)\,;\, |f(\boldsymbol{x})| \leq c\|\boldsymbol{x}\|^{-a}\,. \qquad (2.35)
$$

Then, the following inequality holds for any $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$ *and any* i *that* $1 \leq i \leq n$,

$$
\begin{aligned}
& \Big| \int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t} \;+\; \sigma^2(2\pi\sigma^2)^{\frac{n}{2}} \frac{\partial^2}{\partial x_i^2} g(\boldsymbol{x};\sigma) \Big| \qquad (2.36) \\
\leq \quad & \frac{3}{2\sigma^2}(M+M^*)^2 \|f\| \,\Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n}\Big)^{\frac{1}{2}} \qquad (2.37) \\
& + \frac{3\,c\,n\,\pi^{\frac{n}{2}} M^{*(n-a)}}{2\sigma^2\Gamma(\frac{n}{2}+1)} \Big(\frac{M^2}{a-n} \;+\; 2\frac{MM^*}{a-n-1} \;+\; \frac{M^{*2}}{a-n-2}\Big)\,. \qquad (2.38)
\end{aligned}
$$

Proof

$$
\begin{aligned}
& \Big| \int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t} \;+\; \sigma^2 (2\pi\sigma^2)^{\frac{n}{2}} \frac{\partial^2}{\partial x_i^2} g(\boldsymbol{x};\sigma) \Big| && (2.39)\\
= \;& \Big| \int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t} \;+\; \sigma^2 (2\pi\sigma^2)^{\frac{n}{2}} \frac{\partial^2}{\partial x_i^2} \int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\, d\boldsymbol{t} \Big| \\
= \;& \Big| \int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t} \;+\; \sigma^2 (2\pi\sigma^2)^{\frac{n}{2}} \int_{\mathbb{R}^n} f(\boldsymbol{t}) \big(\frac{(x_i-t_i)^2}{\sigma^4} - \frac{1}{\sigma^2} \big) k(\boldsymbol{x}-\boldsymbol{t};\sigma)\, d\boldsymbol{t} \Big| \\
= \;& \Big| \int_{\mathbb{R}^n} f(\boldsymbol{t}) \Big(1 \;+\; \big(\frac{(x_i-t_i)^2}{\sigma^2} - 1 \big) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big)\, d\boldsymbol{t} \Big| \\
= \;& \Big| \int_{\mathcal{B}(\mathbf{0},M^*)} f(\boldsymbol{t}) \Big(1 \;+\; \big(\frac{(x_i-t_i)^2}{\sigma^2} - 1 \big) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big)\, d\boldsymbol{t} \Big| \\
& + \Big| \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0},M^*)} f(\boldsymbol{t}) \Big(1 \;+\; \big(\frac{(x_i-t_i)^2}{\sigma^2} - 1 \big) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big)\, d\boldsymbol{t} \Big| .
\end{aligned}
$$

We now bound each of these terms separately. For the first term we proceed as follows.

$$
\begin{aligned}
& \Big| \int_{\mathcal{B}(\mathbf{0},M^*)} f(\boldsymbol{t}) \Big(1 \;+\; \big(\frac{(x_i-t_i)^2}{\sigma^2} - 1 \big) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big)\, d\boldsymbol{t} \Big| && (2.40)\\
\leq \;& \Big(\int_{\mathcal{B}(\mathbf{0},M^*)} f^2(\boldsymbol{t})\, d\boldsymbol{t} \Big)^{1/2} \Bigg(\int_{\mathcal{B}(\mathbf{0},M^*)} \Big(1 \;+\; \big(\frac{(x_i-t_i)^2}{\sigma^2} - 1 \big) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big)^2 d\boldsymbol{t} \Bigg)^{1/2} && (2.41)\\
\leq \;& \|f\| \left(\int_{\mathcal{B}(\mathbf{0},M^*)} \Big(1 \;+\; \big(\frac{(x_i-t_i)^2}{\sigma^2} - 1 \big) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big)^2 d\boldsymbol{t} \right)^{1/2} && (2.42)\\
\leq \;& \|f\| \left(\int_{\mathcal{B}(\mathbf{0},M^*)} \Big(\frac{3}{2} \frac{(x_i-t_i)^2}{\sigma^2} \Big)^2 d\boldsymbol{t} \right)^{1/2} && (2.43)\\
= \;& \frac{3}{2\sigma^2} \|f\| \left(\int_{\mathcal{B}(\mathbf{0},M^*)} \Big((x_i-t_i)^2 \Big)^2 d\boldsymbol{t} \right)^{1/2} && (2.44)\\
\leq \;& \frac{3}{2\sigma^2} \|f\| \left(\int_{\mathcal{B}(\mathbf{0},M^*)} \Big((M+M^*)^2 \Big)^2 d\boldsymbol{t} \right)^{1/2} && (2.45)\\
= \;& \frac{3}{2\sigma^2} (M+M^*)^2 \|f\| \Big(\mathrm{Vol}(\mathcal{B}(\mathbf{0},M^*)) \Big)^{\frac{1}{2}} && (2.46)\\
= \;& \frac{3}{2\sigma^2} (M+M^*)^2 \|f\| \Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \Big)^{\frac{1}{2}}, && (2.47)
\end{aligned}
$$

where (2.41) uses Cauchy-Schwartz inequality, and (2.43) uses Proposition

4 (Bound on Hessian's Diagonal of a Gaussian) and (2.45) uses the fact that integration variable $\boldsymbol{t} \in \mathcal{B}(\mathbf{0}, M^*)$ and the assumption that $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$,

We now proceed with the second term as the following:

$$
\begin{aligned}
& \Big| \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} f(\boldsymbol{t}) \Big(1 \; + \; \big(\frac{(x_i - t_k)^2}{\sigma^2} - 1 \big) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big) \, d\boldsymbol{t} \, \Big| && (2.48) \\
\leq \; & \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} |f(\boldsymbol{t})| \quad \Big| 1 \; + \; \big(\frac{(x_i - t_k)^2}{\sigma^2} - 1 \big) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big| \; d\boldsymbol{t} && (2.49) \\
\leq \; & \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} |f(\boldsymbol{t})| \frac{3}{2} \frac{(x_i - t_k)^2}{\sigma^2} \, d\boldsymbol{t} && (2.50) \\
\leq \; & \frac{3}{2\sigma^2} \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} c \|\boldsymbol{t}\|^{-a} (x_i - t_k)^2 \, d\boldsymbol{t} && (2.51) \\
\leq \; & \frac{3}{2\sigma^2} \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} c \|\boldsymbol{t}\|^{-a} (x_i^2 + 2|x_i| \, |t_k| + t_k^2) \, d\boldsymbol{t} && (2.52) \\
\leq \; & \frac{3}{2\sigma^2} \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} c \|\boldsymbol{t}\|^{-a} (M^2 + 2M \, |t_k| + t_k^2) \, d\boldsymbol{t} && (2.53) \\
\leq \; & \frac{3}{2\sigma^2} \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} c \|\boldsymbol{t}\|^{-a} (M^2 + 2M \, \|\boldsymbol{t}\| + \|\boldsymbol{t}\|^2) \, d\boldsymbol{t} && (2.54) \\
= \; & \frac{3c}{2\sigma^2} \int_{M^*}^{\infty} \mathrm{Surf}(\mathcal{S}_{n-1}(\mathbf{0}, r)) \Big(M^2 r^{-a} \; + \; 2M r^{1-a} \; + \; r^{2-a} \Big) \, dr && (2.55) \\
= \; & \frac{3c}{2\sigma^2} \int_{M^*}^{\infty} \frac{n \pi^{\frac{n}{2}} r^{n-1}}{\Gamma(\frac{n}{2} + 1)} \Big(M^2 r^{-a} \; + \; 2M r^{1-a} \; + \; r^{2-a} \Big) \, dr && (2.56) \\
= \; & \frac{3 \, c \, n \, \pi^{\frac{n}{2}} M^{*(n-a)}}{2\sigma^2 \Gamma(\frac{n}{2} + 1)} \Big(\frac{M^2}{a - n} \; + \; 2 \frac{M M^*}{a - n - 1} \; + \; \frac{M^{*2}}{a - n - 2} \Big), && (2.57)
\end{aligned}
$$

where (2.50) uses Proposition 4 (Bound on Hessian's Diagonal of a Gaussian), and (2.51) applies lemma's assumption $\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*) \, ; \, |f(\boldsymbol{x})| \leq c \|\boldsymbol{x}\|^{-a}$. In (2.53) we use lemma's assumption that $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$. Finally, (2.56) uses the fact that the integral of a ***radially symmetric*** function is equivalent to a 1-d integral along the radius of $(n-1)$-dimensional sphere.

Applying the inequalities in (2.47) and (2.57) to (2.39), the it follows that:

$$
\begin{aligned}
& \Big| \int_{\mathbb{R}^n} f(\boldsymbol{x})\, d\boldsymbol{x} \;+\; \sigma^2 (2\pi\sigma^2)^{\frac{n}{2}} \frac{\partial}{\partial x_k^2} g(\boldsymbol{x};\sigma) \Big| && (2.58)\\
\leq \quad & \frac{3}{2\sigma^2}(M+M^*)^2 \|f\| \left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \right)^{\frac{1}{2}} && (2.59)\\
& + \frac{3\, c\, n\, \pi^{\frac{n}{2}} M^{*(n-a)}}{2\sigma^2 \Gamma(\frac{n}{2}+1)} \Big(\frac{M^2}{a-n} \;+\; 2\frac{MM^*}{a-n-1} \;+\; \frac{M^{*2}}{a-n-2} \Big)\,. && (2.60)
\end{aligned}
$$

□

Proposition 6 (Bound on Hessian's Off-Diagonal of a Gaussian) *For any real vectors $\boldsymbol{x}$ and $\boldsymbol{t}$ and any $\sigma > 0$, the following inequality holds.*

$$
0 \leq \frac{|x_i - t_i|\; |x_j - t_j|}{\sigma^2} e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \leq \frac{|x_i - t_i|\; |x_j - t_j|}{\sigma^2} \tag{2.61}
$$

Proof Consider the function $|z|e^{-\frac{z^2}{2}}$ defined for any $z \in \mathbb{R}$. We first obtain an over-estimator for this function that has simpler form. It is easy to check that $|z|$ provides such an over estimator:

$$
\forall z \in \mathbb{R} \quad ; \quad |z| \geq |z| e^{-\frac{z^2}{2}} \geq 0\,. \tag{2.62}
$$

In particular, by choosing $z = (x_i - t_i)/\sigma$ and then again $z = (x_j - t_j)/\sigma$, we can proceed as below:

$$
\begin{aligned}
& \frac{|x_i - t_i|}{\sigma} \geq \frac{|x_i - t_i|}{\sigma} e^{-\frac{(x_i-t_i)^2}{2\sigma^2}} \geq 0 \\
\Rightarrow \quad & \frac{|x_i - t_i|\; |x_j - t_j|}{\sigma^2} \geq \frac{|x_i - t_i|\; |x_j - t_j|}{\sigma^2} e^{-\frac{(x_i-t_i)^2+(x_j-t_j)^2}{2\sigma^2}} \geq 0 \\
\Rightarrow \quad & \frac{|x_i - t_i|\; |x_j - t_j|}{\sigma^2} \geq \frac{|x_i - t_i|\; |x_j - t_j|}{\sigma^2} e^{-\frac{(x_i-t_i)^2+(x_j-t_j)^2}{2\sigma^2}} e^{-\frac{\sum_{i\neq j, i\neq k}(x_i-t_i)^2}{2\sigma^2}} \geq -1 \\
\equiv \quad & \frac{|x_i - t_i|\; |x_j - t_j|}{\sigma^2} \geq \frac{|x_i - t_i|\; |x_j - t_j|}{\sigma^2} e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \geq 0\,. && (2.63)
\end{aligned}
$$

□

Lemma 7 (Convergence of Hessian's Off-Diagonal) *Consider $f : \mathbb{R}^n \to \mathbb{R}$. Suppose there exists for f some $M^* \geq 0$, $c \geq 0$ and integer $a \geq n+3$ with the following property:*

$$\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)\,;\, |f(\boldsymbol{x})| \leq c\|\boldsymbol{x}\|^{-a}\,. \tag{2.64}$$

Then, the following inequality holds for any $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$.

$$\begin{aligned}
& |\,\sigma^2(2\pi\sigma^2)^{\frac{n}{2}} \frac{\partial^2}{\partial x_i x_j} g(\boldsymbol{x};\sigma)\,| && (2.65)\\
\leq\quad & \frac{(M+M^*)^2}{\sigma^2}\|f\|\,\Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\Big)^{\frac{1}{2}} && (2.66)\\
& +\frac{c\,n\,\pi^{\frac{n}{2}}M^{*(n-a)}}{\sigma^2\Gamma(\frac{n}{2}+1)}\Big(\frac{M^2}{a-n}\,+\,2\frac{MM^*}{a-n-1}\,+\,\frac{M^{*2}}{a-n-2}\Big)\,. && (2.67)
\end{aligned}$$

Proof We proceed as below:

$$\begin{aligned}
& |\,\sigma^2(2\pi\sigma^2)^{\frac{n}{2}} \frac{\partial^2}{\partial x_i x_j} g(\boldsymbol{x};\sigma)\,| && (2.68)\\
=\quad & |\,\sigma^2(2\pi\sigma^2)^{\frac{n}{2}} \frac{\partial^2}{\partial x_i x_j} \int_{\mathbb{R}^n} f(\boldsymbol{t})k(\boldsymbol{x}-\boldsymbol{t};\sigma)\,d\boldsymbol{t}\,| && (2.69)\\
=\quad & |\,\sigma^2(2\pi\sigma^2)^{\frac{n}{2}} \int_{\mathbb{R}^n} \frac{(x_i-t_i)(x_j-t_j)}{\sigma^4} f(\boldsymbol{t})k(\boldsymbol{x}-\boldsymbol{t};\sigma)\,d\boldsymbol{t}\,| && (2.70)\\
=\quad & |\int_{\mathcal{B}(\mathbf{0},M^*)} \frac{(x_i-t_i)(x_j-t_j)}{\sigma^2} f(\boldsymbol{t})e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\,d\boldsymbol{t}\,| && (2.71)\\
& +|\int_{\mathbb{R}^n\setminus\mathcal{B}(\mathbf{0},M^*)} \frac{(x_i-t_i)(x_j-t_j)}{\sigma^2} f(\boldsymbol{t})e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\,d\boldsymbol{t}\,|\,. && (2.72)
\end{aligned}$$

We now bound each of these terms separately. For the first term we proceed as follows.

$$
\begin{aligned}
& \Big| \int_{\mathcal{B}(\mathbf{0},M^*)} \frac{(x_i - t_i)(x_j - t_j)}{\sigma^2} f(\boldsymbol{t}) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \, d\boldsymbol{t} \, \Big| && (2.73) \\
\leq \; & \Big(\int_{\mathcal{B}(\mathbf{0},M^*)} f^2(\boldsymbol{t}) \, d\boldsymbol{t} \Big)^{1/2} && (2.74) \\
& \times \Bigg(\int_{\mathcal{B}(\mathbf{0},M^*)} \Big(\frac{(x_i - t_i)(x_j - t_j)}{\sigma^2} e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big)^2 d\boldsymbol{t} \Bigg)^{1/2} && (2.75) \\
\leq \; & \|f\| \Bigg(\int_{\mathcal{B}(\mathbf{0},M^*)} \Big(\frac{(x_i - t_i)(x_j - t_j)}{\sigma^2} e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big)^2 d\boldsymbol{t} \Bigg)^{1/2} && (2.76) \\
\leq \; & \|f\| \Bigg(\int_{\mathcal{B}(\mathbf{0},M^*)} \Big(\frac{|x_i - t_i| \; |x_j - t_j|}{\sigma^2} \Big)^2 d\boldsymbol{t} \Bigg)^{1/2} && (2.77) \\
\leq \; & \frac{1}{\sigma^2} \|f\| \Bigg(\int_{\mathcal{B}(\mathbf{0},M^*)} (M + M^*)^2 \, (M + M^*)^2 \, d\boldsymbol{t} \Bigg)^{1/2} && (2.78) \\
= \; & \frac{(M + M^*)^2}{\sigma^2} \|f\| \, \Big(\, \mathrm{Vol} \left(\mathcal{B}(\mathbf{0}, M^*) \right) \Big)^{\frac{1}{2}} && (2.79) \\
= \; & \frac{(M + M^*)^2}{\sigma^2} \|f\| \, \Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)} M^{*n} \Big)^{\frac{1}{2}}, && (2.80)
\end{aligned}
$$

where (2.74) uses Cauchy-Schwartz inequality, and (2.77) uses Proposition 6 (Bound on Hessian's Off-Diagonal of a Gaussian).

We now proceed with the second term as the following.

$$
\begin{aligned}
& \Big| \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} \frac{(x_i - t_i)(x_j - t_j)}{\sigma^2} f(\boldsymbol{t}) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \, d\boldsymbol{t} \, \Big| \\
\leq & \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} |f(\boldsymbol{t})| \quad \Big| \frac{(x_i - t_i)(x_j - t_j)}{\sigma^2} e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big| \; d\boldsymbol{t} \\
\leq & \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} |f(\boldsymbol{t})| \frac{|x_i - t_i| \; |x_j - t_j|}{\sigma^2} \, d\boldsymbol{t} \qquad (2.81) \\
\leq & \frac{1}{\sigma^2} \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} c \|\boldsymbol{t}\|^{-a} |x_i - t_i| \; |x_j - t_j| \, d\boldsymbol{t} \qquad (2.82) \\
\leq & \frac{1}{\sigma^2} \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} c \|\boldsymbol{t}\|^{-a} (|x_i| + |t_i|)(|x_j| + |t_j|) \, d\boldsymbol{t} \\
= & \frac{1}{\sigma^2} \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} c \|\boldsymbol{t}\|^{-a} (\, |x_i| \, |x_j| \; + \; |x_i| \, |t_j| \; + \; |t_i| \, |x_j| \; + \; |t_i| \, |t_j| \,) \, d\boldsymbol{t} \\
\leq & \frac{1}{\sigma^2} \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} c \|\boldsymbol{t}\|^{-a} (\, |x_i| \, |x_j| \; + \; (|x_i| \; + \; |x_j|) \|\boldsymbol{t}\| \; + \; \|\boldsymbol{t}\|^2 \,) \, d\boldsymbol{t} \\
= & \frac{c}{\sigma^2} \int_{M^*}^{\infty} \mathrm{Surf}(\mathcal{S}_{n-1}(\mathbf{0}, r)) r^{-a} (|x_i| \, |x_j| + (|x_i| + |x_j|) r + r^2) d\boldsymbol{t} \qquad (2.83) \\
= & \frac{c}{\sigma^2} \int_{M^*}^{\infty} \frac{n \, \pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)} r^{n-1} \; r^{-a} (\, |x_i| \, |x_j| \; + \; (|x_i| \; + \; |x_j|) r \; + \; r^2 \,) \, dr \\
\leq & \frac{c \, n \, \pi^{\frac{n}{2}}}{\sigma^2 \Gamma(\frac{n}{2} + 1)} \int_{M^*}^{\infty} r^{n-1-a} (M^2 \; + \; 2rM \; + \; r^2 \,) \, dr \qquad (2.84) \\
= & \frac{c \, n \, \pi^{\frac{n}{2}} M^{*(n-a)}}{\sigma^2 \Gamma(\frac{n}{2} + 1)} \Big(\frac{M^2}{a - n} \; + \; 2 \frac{M M^*}{a - n - 1} \; + \; \frac{M^{*2}}{a - n - 2} \Big) , \qquad (2.85)
\end{aligned}
$$

where (2.81) uses Proposition 6 (Bound on Hessian's Off-Diagonal of a Gaussian), and (2.82) applies lemma's assumption, $\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*) \, ; \, |f(\boldsymbol{x})| \leq c \|\boldsymbol{x}\|^{-a}$. In (2.84) we use the fact that $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$. Finally, (2.83) uses the fact that the integral of a ***radially symmetric*** function is equivalent to a 1-d integral along the radius of $n-1$-dimensional sphere.

Applying the inequalities in (2.80) and (2.85) to (2.68), the following follows.

$$
\begin{aligned}
& |\, \sigma^2(2\pi\sigma^2)^{\frac{n}{2}} \frac{\partial^2}{\partial x_i x_j} g(\boldsymbol{x};\sigma)\,| && (2.86)\\
\leq \quad & \frac{(M+M^*)^2}{\sigma^2}\|f\| \, \Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \Big)^{\frac{1}{2}} && (2.87)\\
& + \frac{c\, n\, \pi^{\frac{n}{2}} M^{*(n-a)}}{\sigma^2 \Gamma(\frac{n}{2}+1)} \Big(\frac{M^2}{a-n} \, + \, 2\frac{MM^*}{a-n-1} \, + \, \frac{M^{*2}}{a-n-2} \Big) . && (2.88)
\end{aligned}
$$

□

Theorem 8 (Convergence of Hessian) *Consider $f : \mathbb{R}^n \to \mathbb{R}$ such that $\|f\|$ is bounded. Suppose there exists for f some $M^* \geq 0$, $c \geq 0$ and integer $a \geq n+3$ with the following property.*

$$
\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)\,;\, |f(\boldsymbol{x})| \leq c\|\boldsymbol{x}\|^{-a}\,. \tag{2.89}
$$

Suppose for any $\epsilon > 0$ and any $M > 0$, σ (as a function of ϵ and M) is chosen large enough in the following sense,

$$
\begin{aligned}
\sigma^2 \quad \geq \quad & \frac{n}{\epsilon}(n+\frac{1}{2}) \Bigg(\|f\|(M^2+M^{*2}) \, \Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \Big)^{\frac{1}{2}} && (2.90)\\
& + \frac{c\, n\, \pi^{\frac{n}{2}} M^{*(n-a)}}{\Gamma(\frac{n}{2}+1)} \Big(\frac{M^2}{a-n} \, + \, 2\frac{MM^*}{a-n-1} \, + \, \frac{M^{*2}}{a-n-2} \Big) \Bigg) && (2.91)
\end{aligned}
$$

Then it is guaranteed to have[3]*,*

$$
\forall \boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)\,;\, \|\boldsymbol{I} \int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t} \, + \, \sigma^2(\sqrt{2\pi}\sigma)^n \nabla^2 g(\boldsymbol{x};\sigma)\|_\infty \leq \epsilon\,. \tag{2.92}
$$

Proof We first upper bound $\|\boldsymbol{I} \int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t} + \sigma^2(\sqrt{2\pi}\sigma)^n \nabla^2 g(\boldsymbol{x};\sigma)\|_\infty$ as follows,

[3]The notation $\|\boldsymbol{A}\|_\infty$ is for the max-norm of the matrix $\boldsymbol{A}$ and is defined as $\|\boldsymbol{A}\|_\infty \triangleq \max_{i,j} |a_{ij}|$.

$$
\begin{aligned}
& \|\boldsymbol{I}\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} \;+\; \sigma^2(\sqrt{2\pi}\sigma)^n\nabla^2 g(\boldsymbol{x};\sigma)\|_\infty \\
\leq\; & \sum_{i=1}^{n} |\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} \;+\; \sigma^2(\sqrt{2\pi}\sigma)^n \frac{\partial^2}{\partial x_i^2} g(\boldsymbol{x};\sigma)\,| \\
& + \sum_{i=1}^{n}\sum_{j\neq i}^{n} |\sigma^2(\sqrt{2\pi}\sigma)^n \frac{\partial^2}{\partial x_i x_j} g(\boldsymbol{x};\sigma)\,| \\
\leq\; & \sum_{i=1}^{n} \frac{3}{2\sigma^2}(M+M^*)^2\|f\| \,\Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\Big)^{\frac{1}{2}} \\
& \quad + \frac{3\,c\,n\,\pi^{\frac{n}{2}}M^{*(n-a)}}{2\sigma^2\Gamma(\frac{n}{2}+1)}\Big(\frac{M^2}{a-n} \;+\; 2\frac{MM^*}{a-n-1} \;+\; \frac{M^{*2}}{a-n-2}\Big) \\
& + \sum_{i=1}^{n}\sum_{j\neq k}^{n} \frac{(M+M^*)^2}{\sigma^2}\|f\| \,\Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\Big)^{\frac{1}{2}} \qquad (2.93) \\
& \quad + \frac{c\,n\,\pi^{\frac{n}{2}}M^{*(n-a)}}{\sigma^2\Gamma(\frac{n}{2}+1)}\Big(\frac{M^2}{a-n} \;+\; 2\frac{MM^*}{a-n-1} \;+\; \frac{M^{*2}}{a-n-2}\Big) \\
\leq\; & \frac{n}{\sigma^2}(\frac{3}{2}+n-1)\Bigg(\|f\|(M^2+M^{*2}) \,\Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\Big)^{\frac{1}{2}} \\
& + \frac{c\,n\,\pi^{\frac{n}{2}}M^{*(n-a)}}{\Gamma(\frac{n}{2}+1)}\Big(\frac{M^2}{a-n} \;+\; 2\frac{MM^*}{a-n-1} \;+\; \frac{M^{*2}}{a-n-2}\Big)\Bigg), \qquad (2.94)
\end{aligned}
$$

where (2.93) uses Lemma 5 (Convergence of Hessian's Diagonals) and (2.93) uses Lemma 7 (Convergence of Hessian's Diagonals). Also in the above inequalities are well-defined due to the assumption that $\|f\| \neq \infty$.

In order to guarantee $\|\boldsymbol{I}\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} + \sigma^2(\sqrt{2\pi}\sigma)^n\nabla^2 g(\boldsymbol{x};\sigma)\|_\infty$ is less than ϵ, it is sufficient to have an upper bound of the former being less than ϵ. That means, the following inequality must hold.

$$\frac{n}{\sigma^2}(\frac{3}{2}+n-1)\Bigg(\|f\|(M^2+M^{*2})\Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\Big)^{\frac{1}{2}} \tag{2.95}$$

$$+\frac{c\,n\,\pi^{\frac{n}{2}}M^{*(n-a)}}{\Gamma(\frac{n}{2}+1)}\Big(\frac{M^2}{a-n}+2\frac{MM^*}{a-n-1}+\frac{M^{*2}}{a-n-2}\Big)\Bigg) \tag{2.96}$$

$$\leq \epsilon \tag{2.97}$$

$$\equiv \quad \frac{n}{\epsilon}(n+\frac{1}{2})\Bigg(\|f\|(M^2+M^{*2})\Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\Big)^{\frac{1}{2}} \tag{2.98}$$

$$+\frac{c\,n\,\pi^{\frac{n}{2}}M^{*(n-a)}}{\Gamma(\frac{n}{2}+1)}\Big(\frac{M^2}{a-n}+2\frac{MM^*}{a-n-1}+\frac{M^{*2}}{a-n-2}\Big)\Bigg) \tag{2.99}$$

$$\leq \sigma^2\,. \tag{2.100}$$

□

Corollary 9 (Asymptotic Convexity for Functions with Rapid Decay) *Consider a continuous function $f:\mathbb{R}^n\to\mathbb{R}$. Suppose there exists for f some $M^*\geq 0$, $c\geq 0$ and integer $a\geq n+3$ with the following property,*

$$\forall \boldsymbol{x}\in\mathbb{R}^n\setminus\mathcal{B}(\boldsymbol{0},M^*)\,;\,|f(\boldsymbol{x})|\leq c\|\boldsymbol{x}\|^{-a}\,. \tag{2.101}$$

If $\int_{\mathbb{R}^n} f(\boldsymbol{x})\,d\boldsymbol{x}<0$ then f is asymptotically strict convex.
More precisely, it will hold that,

$$\begin{aligned}&\forall M>0\,,\,\exists\,\sigma_M^*\,,\,\forall \boldsymbol{x}\in\mathcal{B}(\boldsymbol{0},M)\,,\,\forall\sigma\geq\sigma_M^*\,:\\ &\nabla^2 g(\boldsymbol{x};\sigma)\succ\boldsymbol{O}\,,\end{aligned} \tag{2.102}$$

where σ_M^* is defined as below,

$$\sigma_M^{*\,2}\ \triangleq\ \frac{n^2(n+\frac{1}{2})}{-\int_{\mathbb{R}^n}f(\boldsymbol{t})\,d\boldsymbol{t}}\Bigg(\|f\|(M^2+M^{*2})\Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\Big)^{\frac{1}{2}} \tag{2.103}$$

$$+\frac{cn\pi^{\frac{n}{2}}M^{*(n-a)}}{\Gamma(\frac{n}{2}+1)}\Big(\frac{M^2}{a-n}+2\frac{MM^*}{a-n-1}+\frac{M^{*2}}{a-n-2}\Big)\Bigg)\,. \tag{2.104}$$

Proof Let $\lambda_{\min}(\boldsymbol{A})$ denote the smallest (signed) eigenvalue of the matrix

$\boldsymbol{A}$. We use[4] the inequality $\lambda_{\min}(\boldsymbol{A}+\boldsymbol{B}) \geq \lambda_{\min}(\boldsymbol{A}) + \lambda_{\min}(\boldsymbol{B})$.

$$
\begin{aligned}
& \lambda_{\min}(\sigma^2(\sqrt{2\pi}\sigma)^n \nabla^2 g(\boldsymbol{x};\sigma)) && (2.109)\\
\geq\ & \lambda_{\min}(\boldsymbol{I}\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} + \sigma^2(\sqrt{2\pi}\sigma)^n \nabla^2 g(\boldsymbol{x};\sigma)) && (2.110)\\
& + \lambda_{\min}(-\boldsymbol{I}\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}) && (2.111)\\
\equiv\ & \lambda_{\min}(\sigma^2(\sqrt{2\pi}\sigma)^n \nabla^2 g(\boldsymbol{x};\sigma)) && (2.112)\\
\geq\ & \lambda_{\min}(\boldsymbol{I}\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} \,+\, \sigma^2(\sqrt{2\pi}\sigma)^n \nabla^2 g(\boldsymbol{x};\sigma)) && (2.113)\\
& - \int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} && (2.114)\\
\equiv\ & \lambda_{\min}(\sigma^2(\sqrt{2\pi}\sigma)^n \nabla^2 g(\boldsymbol{x};\sigma)) + \int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} && (2.115)\\
\geq\ & \lambda_{\min}(\boldsymbol{I}\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} \,+\, \sigma^2(\sqrt{2\pi}\sigma)^n \nabla^2 g(\boldsymbol{x};\sigma))\,. && (2.116)
\end{aligned}
$$

On the other hand, it holds between smallest eigenvalue and (entry-wise) max norm that $\lambda_{\min}(\boldsymbol{A}) \geq -\sqrt{\mathrm{tr}(\boldsymbol{A}^T\boldsymbol{A})} = -\sqrt{\sum_{i,j} a_{i,j}^2} \geq -\sqrt{n^2\|\boldsymbol{A}\|_\infty^2} = -n\|\boldsymbol{A}\|_\infty$. Therefore, it follows that,

$$
\begin{aligned}
& \frac{1}{n}\lambda_{\min}\big(\sigma^2(\sqrt{2\pi}\sigma)^n \nabla^2 g(\boldsymbol{x};\sigma)\big) + \frac{1}{n}\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} && (2.117)\\
\geq\ & -\|\boldsymbol{I}\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} \,+\, \sigma^2(\sqrt{2\pi}\sigma)^n \nabla^2 g(\boldsymbol{x};\sigma)\|_\infty\,. && (2.118)
\end{aligned}
$$

Now let's define σ_0 (as a function of ϵ and M) as below,

[4]To derive this, simply use the definition of $\lambda_{\max}$ as follows,

$$
\begin{aligned}
\lambda_{\min}(\boldsymbol{A}+\boldsymbol{B}) &= \inf_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^T(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{v} && (2.105)\\
&= \inf_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v} + \boldsymbol{v}^T\boldsymbol{B}\boldsymbol{v} && (2.106)\\
&\geq \inf_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v} + \inf_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^T\boldsymbol{B}\boldsymbol{v} && (2.107)\\
&= \lambda_{\min}(\boldsymbol{A}) + \lambda_{\min}(\boldsymbol{B})\,. && (2.108)
\end{aligned}
$$

$$\sigma_0{}^2 \triangleq \frac{n}{\epsilon}(n+\frac{1}{2})\Bigg(\|f\|(M^2+M^{*2})\Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\Big)^{\frac{1}{2}} \tag{2.119}$$

$$+\frac{cn\pi^{\frac{n}{2}}M^{*(n-a)}}{\Gamma(\frac{n}{2}+1)}\Big(\frac{M^2}{a-n}+2\frac{MM^*}{a-n-1}+\frac{M^{*2}}{a-n-2}\Big)\Bigg)\,, \tag{2.120}$$

where a and c are determined by the tail decay assumption (2.101). Note that such σ_0 exists because the RHS is finite[5] as $\|f\| \neq \infty$. By Theorem 8 (Convergence of Hessian), it follows that,

$$\forall \boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)\ \forall \sigma > \sigma_0\ ;\ \|\boldsymbol{I}\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} + \sigma^2(\sqrt{2\pi}\sigma)^n\nabla^2 g(\boldsymbol{x};\sigma)\|_\infty \leq \epsilon\,. \tag{2.121}$$

Combining (2.117) and (2.121) leads to the following,

$$\forall \boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)\ \forall \sigma > \sigma_0\ ;\ \frac{1}{n}\lambda_{\min}\big(\sigma^2(\sqrt{2\pi}\sigma)^n\nabla^2 g(\boldsymbol{x};\sigma)\big)+\frac{1}{n}\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} \geq -\epsilon\,. \tag{2.122}$$

Thus, in order to keep $\lambda_{\min}\big(\sigma^2(\sqrt{2\pi}\sigma)^n\nabla^2 g(\boldsymbol{x};\sigma)\big)$ positive within the ball $\mathcal{B}(\mathbf{0}, M)$, it is sufficient to have $-\frac{1}{n}\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}-\epsilon > 0$, or equivalently to have $-\frac{1}{n}\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} > \epsilon$. This can be expressed as follows,

$$\forall \boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)\ \forall \sigma > \sigma_0\ ; \tag{2.123}$$

$$-\frac{1}{n}\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} > \epsilon \Rightarrow \frac{1}{n}\sigma^2(\sqrt{2\pi}\sigma)^n\nabla^2 g(\boldsymbol{x};\sigma) \succ \boldsymbol{O}\,. \tag{2.124}$$

Now ϵ can be eliminated in the above condition, using the equality in (2.119). Using that, and the facts that $\sigma > 0$ and the assumption that $-\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} > 0$, we obtain following,

[5] In one hand, f has bounded integral outside of the ball $\mathcal{B}(\mathbf{0}, M^*)$ due to the decay assumption in this region. On the other hand, f is continuous within the (finite-volume) ball $\mathcal{B}(\mathbf{0}, M^*)$. Hence, it has bounded integral within this ball. Consequently, the sum the outside and inside integrals has to be bounded.

$$\forall \boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)\ \forall \sigma > \sigma_0\ ; \tag{2.125}$$

$$\sigma_0{}^2 > \frac{n^2(n+\frac{1}{2})}{-\int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t}} \Bigg(\|f\|(M^2 + M^{*2}) \Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \Big)^{\frac{1}{2}} \tag{2.126}$$

$$+ \frac{c\, n\, \pi^{\frac{n}{2}} M^{*(n-a)}}{\Gamma(\frac{n}{2}+1)} \Big(\frac{M^2}{a-n} + 2\frac{MM^*}{a-n-1} + \frac{M^{*2}}{a-n-2} \Big) \Bigg) \tag{2.127}$$

$$\Rightarrow \nabla^2 g(\boldsymbol{x}; \sigma) \succ \boldsymbol{O}\,. \tag{2.128}$$

For clarity, let's define σ_M^* as following,

$$\sigma_M^*{}^2 \triangleq \frac{n^2(n+\frac{1}{2})}{-\int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t}} \Bigg(\|f\|(M^2 + M^{*2}) \Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \Big)^{\frac{1}{2}} \tag{2.129}$$

$$+ \frac{c\, n\, \pi^{\frac{n}{2}} M^{*(n-a)}}{\Gamma(\frac{n}{2}+1)} \Big(\frac{M^2}{a-n} + 2\frac{MM^*}{a-n-1} + \frac{M^{*2}}{a-n-2} \Big) \Bigg) \tag{2.130}$$

Then the expression (2.125), due to the fact that we did not make any assumption on M except that $M > 0$, can be expressed more precisely as follows,

$$\forall M > 0\,,\, \exists\, \sigma_M^*\,,\, \forall \boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)\,,\, \forall \sigma \geq \sigma_M^*\ : \\ \nabla^2 g(\boldsymbol{x}; \sigma) \succ \boldsymbol{O}\,. \tag{2.131}$$

By Proposition 2, the above condition is equivalent to the definition of asymptotic strict convexity of f.

□

2.6.2 Derivative Free Results on Asymptotic Minimizer

Main Result (Corollary 13) *Consider $f : \mathbb{R}^n \to \mathbb{R}$. Suppose there is an origin-centered ball out of which f decays like $\|\boldsymbol{x}\|^{-n-4}$ or faster. Then for any origin-centered ball, there always exists some σ that can make $\frac{\int_{\mathbb{R}^n} \boldsymbol{t}\, f(\boldsymbol{t})\, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})\, d\boldsymbol{t}}$ arbitrary close to the stationary point of $g(\boldsymbol{x}; \sigma)$.*

Lemma 10 (Zeroth Moment Convergence) *Consider $f : \mathbb{R}^n \to \mathbb{R}$. Suppose there exists for f some $M^* \geq 0$, $c \geq 0$ and integer $a \geq n+3$ with the following property.*

$$\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)\,;\, |f(\boldsymbol{x})| \leq c\|\boldsymbol{x}\|^{-a}\,. \tag{2.132}$$

Then for any $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$ the following inequality holds.

$$\begin{aligned} & |\int_{\mathbb{R}^n} f(\boldsymbol{t})\Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\Big)\,d\boldsymbol{t}\,| && (2.133)\\ \leq\ & \frac{(M+M^*)^2}{2\sigma^2}\|f\|\Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\Big)^{\frac{1}{2}} && (2.134)\\ & + \frac{c\,n\pi^{\frac{n}{2}}M^{*n-a}}{2\sigma^2\,\Gamma(\frac{n}{2}+1)}\Big(\frac{M^2}{a-n} + 2\frac{MM^*}{a-n-1} + \frac{M^{*2}}{a-n-2}\Big)\,. && (2.135) \end{aligned}$$

Proof

$$\begin{aligned} & |\int_{\mathbb{R}^n} f(\boldsymbol{t})\Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\Big)\,d\boldsymbol{t}\,| && (2.136)\\ =\ & |\int_{\mathcal{B}(\mathbf{0},M^*)} f(\boldsymbol{t})\Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\Big)\,d\boldsymbol{t}\,| && (2.137)\\ & + |\int_{\mathbb{R}^n\setminus\mathcal{B}(\mathbf{0},M^*)} f(\boldsymbol{t})\Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\Big)\,d\boldsymbol{t}\,|\,. && (2.138) \end{aligned}$$

We now bound each of these terms separately. Starting from the first term in (2.137), we proceed as below,

$$
\begin{aligned}
& \Big| \int_{\mathcal{B}(\mathbf{0},M^*)} f(\boldsymbol{t}) \Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big) \, d\boldsymbol{t} \, \Big| && (2.139) \\
\leq \; & \|f\| \left(\int_{\mathcal{B}(\mathbf{0},M^*)} \Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big)^2 d\boldsymbol{t} \right)^{\frac{1}{2}} && (2.140) \\
\leq \; & \|f\| \left(\int_{\mathcal{B}(\mathbf{0},M^*)} \Big(\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2} \Big)^2 d\boldsymbol{t} \right)^{\frac{1}{2}} && (2.141) \\
\leq \; & \|f\| \left(\int_{\mathcal{B}(\mathbf{0},M^*)} \Big(\frac{(M+M^*)^2}{2\sigma^2} \Big)^2 d\boldsymbol{t} \right)^{\frac{1}{2}} && (2.142) \\
= \; & \frac{(M+M^*)^2}{2\sigma^2} \|f\| \left(\text{Vol}\Big(\mathcal{B}(\mathbf{0}, M^*) \Big) \right)^{\frac{1}{2}} && (2.143) \\
= \; & \frac{(M+M^*)^2}{2\sigma^2} \|f\| \Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \Big)^{\frac{1}{2}}, && (2.144)
\end{aligned}
$$

where (2.140) uses Cauchy-Schwartz inequality and (2.141) uses the fact that $\forall y \in \mathbb{R} \; ; \; 1 - e^{-y^2} \leq y^2$. In (2.142) we use lemma's assumption that $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$ and that the integration domain is $\boldsymbol{t} \in \mathcal{B}(\mathbf{0}, M^*)$.

Now we upper bound the second term in (2.137) as below,

$$
\begin{aligned}
& \Big| \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0},M^*)} f(\boldsymbol{t}) \Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big) \, d\boldsymbol{t} \, \Big| && (2.145) \\
\leq & \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0},M^*)} |\, f(\boldsymbol{t}) \,| \quad |\, 1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \,| \, d\boldsymbol{t} && (2.146) \\
\leq & \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0},M^*)} |\, f(\boldsymbol{t}) \,| \quad \frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2} \, d\boldsymbol{t} && (2.147) \\
\leq & \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0},M^*)} |\, f(\boldsymbol{t}) \,| \quad \frac{\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{x}\| \, \|\boldsymbol{t}\| + \|\boldsymbol{t}\|^2}{2\sigma^2} \, d\boldsymbol{t} && (2.148) \\
\leq & \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0},M^*)} c \, \|\boldsymbol{t}\|^{-a} \, \frac{\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{x}\| \, \|\boldsymbol{t}\| + \|\boldsymbol{t}\|^2}{2\sigma^2} \, d\boldsymbol{t} && (2.149) \\
\leq & \frac{c}{2\sigma^2} \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0},M^*)} M^2 \|\boldsymbol{t}\|^{-a} \, + \, 2M \|\boldsymbol{t}\|^{1-a} \, + \, \|\boldsymbol{t}\|^{2-a} \, d\boldsymbol{t} && (2.150) \\
= & \frac{c}{2\sigma^2} \int_{M^*}^{\infty} \mathrm{Surf}(\mathcal{S}_{n-1}(\mathbf{0}, r)) \Big(M^2 r^{-a} \, + \, 2M r^{1-a} \, + \, r^{2-a} \Big) \, dr && (2.151) \\
= & \frac{c}{2\sigma^2} \int_{M^*}^{\infty} \frac{n\pi^{\frac{n}{2}} r^{n-1}}{\Gamma(\frac{n}{2}+1)} \Big(M^2 r^{-a} \, + \, 2M r^{1-a} \, + \, r^{2-a} \Big) \, dr && (2.152) \\
= & \frac{c \, n\pi^{\frac{n}{2}} M^{*n-a}}{2\sigma^2 \, \Gamma(\frac{n}{2}+1)} \Big(\frac{M^2}{a-n} \, + \, 2 \frac{MM^*}{a-n-1} \, + \, \frac{M^{*2}}{a-n-2} \Big) . && (2.153)
\end{aligned}
$$

Applying (2.144) and (2.153) to (2.137) implies the following inequality,

$$
\begin{aligned}
& \Big| \int_{\mathbb{R}^n} f(\boldsymbol{t}) \Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big) \, d\boldsymbol{t} \, \Big| && (2.154) \\
\leq & \frac{(M+M^*)^2}{2\sigma^2} \|f\| \Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \Big)^{\frac{1}{2}} && (2.155) \\
& + \frac{c \, n\pi^{\frac{n}{2}} M^{*n-a}}{2\sigma^2 \, \Gamma(\frac{n}{2}+1)} \Big(\frac{M^2}{a-n} \, + \, 2 \frac{MM^*}{a-n-1} \, + \, \frac{M^{*2}}{a-n-2} \Big) . && (2.156)
\end{aligned}
$$

□

Lemma 11 (First Moment Convergence) *Consider $f : \mathbb{R}^n \to \mathbb{R}$. Suppose there exists for f some $M^* \geq 0$, $c \geq 0$ and integer $a \geq n+4$ with the following property.*

$$
\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*) \, ; \, |f(\boldsymbol{x})| \leq c \|\boldsymbol{x}\|^{-a} . \tag{2.157}
$$

Define $h_i(\boldsymbol{t}) \triangleq t_i f(\boldsymbol{t})$. Then for any $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$, and any $i = 1, 2, \cdots, n$,

the following inequality holds.

$$
\begin{aligned}
& \Big| \int_{\mathbb{R}^n} h_i(\boldsymbol{t}) \Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big) \, d\boldsymbol{t} \, \Big| && (2.158) \\
\leq \; & \frac{(M+M^*)^2}{2\sigma^2} \|h_i\| \Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \Big)^{\frac{1}{2}} && (2.159) \\
& \frac{c\, n \pi^{\frac{n}{2}} M^{*n+1-a}}{2\sigma^2 \, \Gamma(\frac{n}{2}+1)} \Big(\frac{M^2}{a-n-1} + 2\frac{MM^*}{a-n-2} + \frac{M^{*2}}{a-n-3} \Big) && (2.160)
\end{aligned}
$$

Proof

$$
\begin{aligned}
& \Big| \int_{\mathbb{R}^n} h_i(\boldsymbol{t}) \Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big) \, d\boldsymbol{t} \, \Big| && (2.161) \\
= \; & \Big| \int_{\mathcal{B}(\mathbf{0},M^*)} h_i(\boldsymbol{t}) \Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big) \, d\boldsymbol{t} \, \Big| && (2.162) \\
& + \Big| \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0},M^*)} h_i(\boldsymbol{t}) \Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big) \, d\boldsymbol{t} \, \Big| \, . && (2.163)
\end{aligned}
$$

We now bound each of these terms separately. Starting from the first term in (2.162), we proceed as below.

$$
\begin{aligned}
& \Big| \int_{\mathcal{B}(\mathbf{0},M^*)} h_i(\boldsymbol{t}) \Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big) \, d\boldsymbol{t} \, \Big| && (2.164) \\
\leq \; & \|h_i\| \left(\int_{\mathcal{B}(\mathbf{0},M^*)} \Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big)^2 d\boldsymbol{t} \right)^{\frac{1}{2}} && (2.165) \\
\leq \; & \|h_i\| \left(\int_{\mathcal{B}(\mathbf{0},M^*)} \Big(\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2} \Big)^2 d\boldsymbol{t} \right)^{\frac{1}{2}} && (2.166) \\
\leq \; & \|h_i\| \left(\int_{\mathcal{B}(\mathbf{0},M^*)} \Big(\frac{(M+M^*)^2}{2\sigma^2} \Big)^2 d\boldsymbol{t} \right)^{\frac{1}{2}} && (2.167) \\
= \; & \frac{(M+M^*)^2}{2\sigma^2} \|h_i\| \left(\mathrm{Vol} \Big(\mathcal{B}(\mathbf{0}, M^*) \Big) \right)^{\frac{1}{2}} && (2.168) \\
= \; & \frac{(M+M^*)^2}{2\sigma^2} \|h_i\| \Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \Big)^{\frac{1}{2}} , && (2.169)
\end{aligned}
$$

where (2.165) uses Cauchy-Schwartz inequality and (2.166) uses the fact that $\forall y \in \mathbb{R} \; ; \; 1 - e^{-y^2} \leq y^2$. In (2.167) we use lemma's assumption that

$\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$ and that the integration domain implies $\boldsymbol{t} \in \mathcal{B}(\mathbf{0}, M^*)$.

Now we upper bound the second term in (2.162) as below.

$$
\begin{aligned}
& \Big| \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} h_i(\boldsymbol{t}) \Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big) \, d\boldsymbol{t} \, \Big| && (2.170) \\
\leq & \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} | \, h_i(\boldsymbol{t}) \, | \quad | \, 1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \, | \, d\boldsymbol{t} && (2.171) \\
\leq & \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} | \, h_i(\boldsymbol{t}) \, | \quad \frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2} \, d\boldsymbol{t} && (2.172) \\
\leq & \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} | \, h_i(\boldsymbol{t}) \, | \quad \frac{\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{x}\| \, \|\boldsymbol{t}\| + \|\boldsymbol{t}\|^2}{2\sigma^2} \, d\boldsymbol{t} && (2.173) \\
\leq & \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} | \, h_i(\boldsymbol{t}) \, | \quad \frac{M^2 + 2M \, \|\boldsymbol{t}\| + \|\boldsymbol{t}\|^2}{2\sigma^2} \, d\boldsymbol{t} && (2.174) \\
\leq & \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} \| \, \boldsymbol{t} \, \| \, | \, f(\boldsymbol{t}) \, | \quad \frac{M^2 + 2M \, \|\boldsymbol{t}\| + \|\boldsymbol{t}\|^2}{2\sigma^2} \, d\boldsymbol{t} && (2.175) \\
\leq & \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} c \, \| \, \boldsymbol{t} \, \| \, \|\boldsymbol{t}\|^{-a} \, \frac{M^2 + 2M \, \|\boldsymbol{t}\| + \|\boldsymbol{t}\|^2}{2\sigma^2} \, d\boldsymbol{t} && (2.176) \\
= & \frac{c}{2\sigma^2} \int_{\mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)} M^2 \|\boldsymbol{t}\|^{1-a} \, + \, 2M \|\boldsymbol{t}\|^{2-a} \, + \, \|\boldsymbol{t}\|^{3-a} \, d\boldsymbol{t} && (2.177) \\
= & \frac{c}{2\sigma^2} \int_{M^*}^{\infty} \mathrm{Surf}(\mathcal{S}_{n-1}(\mathbf{0}, r)) \Big(M^2 r^{1-a} \, + \, 2M r^{2-a} \, + \, r^{3-a} \Big) \, dr && (2.178) \\
= & \frac{c}{2\sigma^2} \int_{M^*}^{\infty} \frac{n \pi^{\frac{n}{2}} r^{n-1}}{\Gamma(\frac{n}{2}+1)} \Big(M^2 r^{1-a} \, + \, 2M r^{2-a} \, + \, r^{3-a} \Big) \, dr && (2.179) \\
= & \frac{c \, n \pi^{\frac{n}{2}} M^{*n+1-a}}{2\sigma^2 \, \Gamma(\frac{n}{2}+1)} \Big(\frac{M^2}{a-n-1} \, + \, 2 \frac{M M^*}{a-n-2} \, + \, \frac{M^{*2}}{a-n-3} \Big) . && (2.180)
\end{aligned}
$$

Applying (2.169) and (2.180) to (2.162) implies the following inequality,

$$
\begin{aligned}
& \Big| \int_{\mathbb{R}^n} h_i(\boldsymbol{t}) \Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}} \Big) \, d\boldsymbol{t} \, \Big| && (2.181) \\
\leq & \frac{(M+M^*)^2}{2\sigma^2} \|h_i\| \Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \Big)^{\frac{1}{2}} && (2.182) \\
& \frac{c \, n \pi^{\frac{n}{2}} M^{*n+1-a}}{2\sigma^2 \, \Gamma(\frac{n}{2}+1)} \Big(\frac{M^2}{a-n-1} \, + \, 2 \frac{M M^*}{a-n-2} \, + \, \frac{M^{*2}}{a-n-3} \Big) && (2.183)
\end{aligned}
$$

□

Theorem 12 *Consider a continuous function $f : \mathbb{R}^n \to \mathbb{R}$. Suppose there*

exists for f some $M^* \geq 0$, $c \geq 0$ and integer $a \geq n+4$ with the following property,

$$\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\mathbf{0}, M^*)\,;\, |f(\boldsymbol{x})| \leq c\|\boldsymbol{x}\|^{-a}\,. \tag{2.184}$$

Define r and u for brevity as the following,

$$r \triangleq \frac{M^2}{a-n} + 2\frac{MM^*}{a-n-1} + \frac{M^{*2}}{a-n-2} \tag{2.185}$$

$$u \triangleq \frac{M^2}{a-n-1} + 2\frac{MM^*}{a-n-2} + \frac{M^{*2}}{a-n-3}\,. \tag{2.186}$$

Then, for any $\epsilon > 0$ and any $i = 1, 2, \cdots, n$, there always exists some $\sigma^* > 0$ that satisfies the following inequality.

$$\begin{aligned}
&\sigma^{*2} \\
\geq\ & \max\{\frac{(M+M^*)^2\|f\|\left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\right)^{\frac{1}{2}} + \frac{cn\pi^{\frac{n}{2}}M^{*n-a}}{\Gamma(\frac{n}{2}+1)}r}{|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}|} \\
&, \frac{|\int_{\mathbb{R}^n} t_i f(\boldsymbol{t})\,d\boldsymbol{t}|}{|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}|}\frac{(M+M^*)^2\|f\|\left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\right)^{\frac{1}{2}} + \frac{cn\pi^{\frac{n}{2}}M^{*n-a}}{\Gamma(\frac{n}{2}+1)}r}{\epsilon|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}|} \\
&+\frac{(M+M^*)^2\|t_i f\|\left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\right)^{\frac{1}{2}} + \frac{cn\pi^{\frac{n}{2}}M^{*n+1-a}}{\Gamma(\frac{n}{2}+1)}u}{\epsilon|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}|}\}\,.
\end{aligned} \tag{2.187}$$

In addition, for any $\sigma \geq \sigma^*$, and for any $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$ the following inequality holds.

$$|\frac{\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}}| \leq \epsilon\,. \tag{2.188}$$

Proof The theorem seeks to bound the following quantity.

$$|\frac{\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}}|\,. \tag{2.189}$$

Since $\sigma > 0$, we have the following identity.

$$\frac{\int_{\mathbb{R}^n} t_i k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}} \tag{2.190}$$

$$= \frac{\int_{\mathbb{R}^n} (\sqrt{2\pi}\sigma)^n f(\boldsymbol{t})t_i k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} (\sqrt{2\pi}\sigma)^n f(\boldsymbol{t})k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}} \tag{2.191}$$

$$= \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\,d\boldsymbol{t}}\,. \tag{2.192}$$

Therefore, the task of bounding (2.189) can be equivalently expressed as bounding the following,

$$\left|\frac{\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\,d\boldsymbol{t}}\right|. \tag{2.193}$$

We first must ensure that this task is *well-defined*, i.e. the denominators are non-zero. We already know from theorem's assumptions that $\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} \neq 0$. Therefore, we just need to ensure $\int_{\mathbb{R}^n} f(\boldsymbol{t})e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\,d\boldsymbol{t} \neq 0$. To do that, we use the following fact[6],

$$\forall (a,b) \in \mathbb{R}-\{0\}\times\mathbb{R}\;;\; |a| \geq 2|a-b| \Rightarrow 2|b| \geq |a|\,. \tag{2.197}$$

Therefore, in order to have $|\int_{\mathbb{R}^n} f(\boldsymbol{t})e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\,d\boldsymbol{t}| \geq \frac{1}{2}|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}|$, it is sufficient to satisfy the following inequality,

[6]In order to see that, we start from $|a| \geq 2|a-b|$ and show that it implies $2|b| \geq |a|$. Observe that $2|a-b| \leq |a|$ can be equivalently expressed as $4(a-b)^2 \leq a^2$, and thus as $3a^2+4b^2-8ab \leq 0$. Since $a \neq 0$ by definition, the LHS is a ***quadratic*** form in a. Therefore, it is nonpositive if and only if a is between the roots of the quadratic. It is easy to derive these roots, which are $\frac{1}{3}(4b \pm 2|b|)$. Hence, we just proved that,

$$|a| \geq 2|a-b| \;\equiv\; \frac{1}{3}(4b-2|b|) \leq a \leq \frac{1}{3}(4b+2|b|)\,. \tag{2.194}$$

In one hand, when $b \geq 0$, $\frac{1}{3}(4b-2|b|) \leq a \leq \frac{1}{3}(4b+2|b|)$ implies that $a \leq 2b$. On the other hand, when $b \leq 0$, $\frac{1}{3}(4b-2|b|) \leq a \leq \frac{1}{3}(4b+2|b|)$ implies that $2b \leq a$. Therefore, for any choice of b, it follows that,

$$\frac{1}{3}(4b-2|b|) \leq a \leq \frac{1}{3}(4b+2|b|) \Rightarrow |a| \leq 2|b|\,, \tag{2.195}$$

which due to the proved equivalence can be written as the following,

$$|a| \geq 2|a-b| \Rightarrow |a| \leq 2|b|\,. \tag{2.196}$$

$$\frac{1}{2}\,|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}\,| \geq |\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} - \int_{\mathbb{R}^n} f(\boldsymbol{t}) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\,d\boldsymbol{t}| \quad (2.198)$$

$$\Rightarrow \quad \int_{\mathbb{R}^n} f(\boldsymbol{t}) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\,d\boldsymbol{t}| \geq \frac{1}{2}\,|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}\,|\,. \quad (2.199)$$

On the other hand, by lemma 10 we have,

$$|\int_{\mathbb{R}^n} f(\boldsymbol{t})\Big(1 - e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\Big)\,d\boldsymbol{t}\,| \quad (2.200)$$

$$\leq \quad \frac{(M+M^*)^2}{2\sigma^2}\|f\|\Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\Big)^{\frac{1}{2}} \quad (2.201)$$

$$+\frac{c\,n\pi^{\frac{n}{2}}M^{*n-a}}{2\sigma^2\,\Gamma(\frac{n}{2}+1)}r\,. \quad (2.202)$$

Plugging this upper bound into (2.198) gives the following sufficient condition for guaranteeing $|\int_{\mathbb{R}^n} f(\boldsymbol{t})e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\,d\boldsymbol{t}\,| \geq \frac{1}{2}\,|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}\,|$.

$$\frac{1}{2}\,|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}\,| \geq \frac{(M+M^*)^2}{2\sigma^2}\|f\|\Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\Big)^{\frac{1}{2}} + \frac{c\,n\pi^{\frac{n}{2}}M^{*n-a}}{2\sigma^2\,\Gamma(\frac{n}{2}+1)}r$$

$$\Rightarrow \quad \sigma \geq \sigma^* \quad (2.203)$$

$$\sigma^{*2} = \frac{(M+M^*)^2\|f\|\Big(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\Big)^{\frac{1}{2}} + \frac{c\,n\pi^{\frac{n}{2}}M^{*n-a}}{\Gamma(\frac{n}{2}+1)}r}{|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}|}\,.$$

It is easy to check that such $\sigma^* > 0$ always exists, because (2.203) is always bounded. Specifically, observe that the assumptions require $\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t} \neq 0$. Also $\|f\| \neq \infty$ due to the tail decay rate condition and continuity of f. Finally, M and M^* cannot be ∞ because they are real numbers. Consequently, $\int_{\mathbb{R}^n} f(\boldsymbol{t})e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\,d\boldsymbol{t} \geq |\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}\,|$ due to (2.197) and since by theorem's assumption we know $|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}\,| \neq 0$, it follows that $\int_{\mathbb{R}^n} f(\boldsymbol{t})e^{-\frac{\|\boldsymbol{x}-\boldsymbol{t}\|^2}{2\sigma^2}}\,d\boldsymbol{t} \neq 0$ as well.

Now that we know how to make (2.189) well-defined, we can proceed by finding an upper bound for it. That is, upper bounding the following,

$$|\frac{\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}}|\,.$$

We start from the fact that $\frac{a}{b}-\frac{c}{d} = a\frac{d-b}{bd}+\frac{a-c}{d}$ and thus we have the following,

$$|\frac{a}{b}-\frac{c}{d}| \leq \frac{|a|}{|b|}\frac{|d-b|}{|d|}+\frac{|a-c|}{|d|}\,. \tag{2.204}$$

In fact, the above inequality implies a more useful one as below.

$$|\frac{a}{b}-\frac{c}{d}| \leq \frac{|a|}{|b|}\frac{\mathrm{UB}(|d-b|)}{\mathrm{LB}(|d|)}+\frac{\mathrm{UB}(|a-c|)}{\mathrm{LB}(|d|)}\,, \tag{2.205}$$

where UB(.) and LB(.) mean any upper bound and lower bound. Using this inequality, we can continue bounding (2.204) as the following,

$$\begin{aligned}
& |\frac{\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}}| \\
\leq\ & \frac{|\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i\,d\boldsymbol{t}|}{|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}|}\frac{\mathrm{UB}(|\int_{\mathbb{R}^n} f(\boldsymbol{t})k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}-\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}|)}{\mathrm{LB}(|\int_{\mathbb{R}^n} f(\boldsymbol{t})k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}|)} \\
& +\frac{\mathrm{UB}(|\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i\,d\boldsymbol{t}-\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}|)}{\mathrm{LB}(|\int_{\mathbb{R}^n} f(\boldsymbol{t})k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}|)} \\
=\ & \frac{|\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i\,d\boldsymbol{t}|}{|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}|}\frac{\frac{(M+M^*)^2}{2\sigma^2}\|f|\left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\right)^{\frac{1}{2}}+\frac{cn\pi^{\frac{n}{2}}M^{*n-a}}{2\sigma^2\,\Gamma(\frac{n}{2}+1)}r}{\frac{1}{2}|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}|} \\
& +\frac{\frac{(M+M^*)^2}{2\sigma^2}\|h_i\|\left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\right)^{\frac{1}{2}}+\frac{cn\pi^{\frac{n}{2}}M^{*n+1-a}}{2\sigma^2\,\Gamma(\frac{n}{2}+1)}u}{\frac{1}{2}|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}|}\,.
\end{aligned} \tag{2.206}$$

Thus, in order to have $|\frac{\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t})t_i k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t})k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}}| \leq \epsilon$, it is sufficient to keep its upper bound in (2.206) less than ϵ as shown below:

$$
\begin{aligned}
&\frac{|\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i \, d\boldsymbol{t}|}{|\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}|} \frac{\frac{(M+M^*)^2}{2\sigma^2} \|f\| \left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \right)^{\frac{1}{2}} + \frac{c\, n \pi^{\frac{n}{2}} M^{*n-a}}{2\sigma^2 \, \Gamma(\frac{n}{2}+1)} r}{\frac{1}{2} |\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}|} \\
&+ \frac{\frac{(M+M^*)^2}{2\sigma^2} \|h_i\| \left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \right)^{\frac{1}{2}} + \frac{c\, n \pi^{\frac{n}{2}} M^{*n+1-a}}{2\sigma^2 \, \Gamma(\frac{n}{2}+1)} u}{\frac{1}{2} |\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}|} \\
&\leq \epsilon \\
\Rightarrow &\left| \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}} \right| \leq \epsilon \,,
\end{aligned}
$$

which can equivalently be written as below,

$$
\begin{aligned}
&\frac{|\int_{\mathbb{R}^n} t_i f(\boldsymbol{t}) \, d\boldsymbol{t}|}{|\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}|} \frac{\frac{(M+M^*)^2}{2\epsilon} \|f\| \left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \right)^{\frac{1}{2}} + \frac{c\, n \pi^{\frac{n}{2}} M^{*n-a}}{2\epsilon \, \Gamma(\frac{n}{2}+1)} r}{\frac{1}{2} |\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}|} \\
&+ \frac{\frac{(M+M^*)^2}{2\epsilon} \|h_i\| \left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} M^{*n} \right)^{\frac{1}{2}} + \frac{c\, n \pi^{\frac{n}{2}} M^{*n+1-a}}{2\epsilon \, \Gamma(\frac{n}{2}+1)} u}{\frac{1}{2} |\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}|} \\
&\leq \sigma^2 \\
\Rightarrow &\left| \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}} \right| \leq \epsilon \,,
\end{aligned}
\tag{2.207}
$$

with σ^* being the σ at the equality. It is easy to check that such $\sigma^* > 0$ always exists, because the LHS of (2.207) is always bounded. Specifically, observe that $\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t} \neq 0$ due to theorem's assumptions, and that $\|f\| < \infty$ and $\|h_i\| < \infty$ due to decay rate[7] property of f.

Remember, however, we earlier had an additional constraint on σ^* back in (2.203). In order for σ^* to jointly satisfy (2.203) and (2.203), we can choose

[7]If there exists for continuous f some $M^* \geq 0$, $c \geq 0$ and integer $a \geq n+2$ with the following property.

$$
\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{B}(\boldsymbol{0}, M^*) \,;\, |f(\boldsymbol{x})| \leq c \|\boldsymbol{x}\|^{-a} \,,
$$

then $\|f\| < \infty$ and $\|h_i\| < \infty$.

it as below,

$$
\begin{aligned}
& \sigma^* \\
= \; & \max\{\frac{(M+M^*)^2\|f\|\left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}\left(M^{*n}\right)^{\frac{1}{2}}+\frac{c\,n\pi^{\frac{n}{2}}M^{*n-a}}{\Gamma(\frac{n}{2}+1)}r\right)}{|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}|} \\
& , \frac{|\int_{\mathbb{R}^n} t_i f(\boldsymbol{t})\,d\boldsymbol{t}|}{|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}|}\frac{(M+M^*)^2\|f\|\left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\right)^{\frac{1}{2}}+\frac{c\,n\pi^{\frac{n}{2}}M^{*n-a}}{\Gamma(\frac{n}{2}+1)}r}{\epsilon|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}|} \\
& +\frac{(M+M^*)^2\|h_i\|\left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\right)^{\frac{1}{2}}+\frac{c\,n\pi^{\frac{n}{2}}M^{*n+1-a}}{\Gamma(\frac{n}{2}+1)}u}{\epsilon|\int_{\mathbb{R}^n} f(\boldsymbol{t})\,d\boldsymbol{t}|}\}\,.
\end{aligned}
$$

□

Corollary 13 *Consider $f:\mathbb{R}^n\to\mathbb{R}$. Suppose there exists for f some $M^*\geq 0$, $c\geq 0$ and integer $a\geq n+4$ with the following property.*

$$\forall \boldsymbol{x}\in\mathbb{R}^n\setminus\mathcal{B}(\mathbf{0},M^*)\,;\,|f(\boldsymbol{x})|\leq c\|\boldsymbol{x}\|^{-a}\,. \tag{2.208}$$

Let $\boldsymbol{x}^*_\sigma$ denote a stationary point of $g(\boldsymbol{x};\sigma)$, that is $\nabla g(\boldsymbol{x}^*_\sigma;\sigma)=\mathbf{0}$. Then, for any $\epsilon>0$ and any $M\geq 0$, there always exists some (large enough) $\sigma>0$ (which depends on ϵ and M) that can make $\|\frac{\int_{\mathbb{R}^n}\boldsymbol{t}\,f(\boldsymbol{t})\,d\boldsymbol{t}}{\int_{\mathbb{R}^n}f(\boldsymbol{t})\,d\boldsymbol{t}}-\boldsymbol{x}^*_\sigma\|_\infty$ arbitrarily small.

Proof Define r and u for brevity as the following,

$$r \;\triangleq\; \frac{M^2}{a-n}+2\frac{MM^*}{a-n-1}+\frac{M^{*2}}{a-n-2} \tag{2.209}$$

$$u \;\triangleq\; \frac{M^2}{a-n-1}+2\frac{MM^*}{a-n-2}+\frac{M^{*2}}{a-n-3}\,. \tag{2.210}$$

The assumption on f allows application of Theorem 12. Therefore, from that theorem it follows that, for any $\epsilon>0$, and any $i=1,2,\cdots,n$, there "always exists" some $\sigma^*>0$ that satisfies the following inequality.

$$
\begin{aligned}
&\sigma^{*2} \geq \\
&\max\{\frac{(M+M^*)^2\|f\|\left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\right)^{\frac{1}{2}}+\frac{cn\pi^{\frac{n}{2}}M^{*n-a}}{\Gamma(\frac{n}{2}+1)}r}{|\int_{\mathbb{R}^n}f(\boldsymbol{t})\,d\boldsymbol{t}|} \\
&,\frac{|\int_{\mathbb{R}^n}t_if(\boldsymbol{t})\,d\boldsymbol{t}|}{|\int_{\mathbb{R}^n}f(\boldsymbol{t})\,d\boldsymbol{t}|}\frac{(M+M^*)^2\|f\|\left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\right)^{\frac{1}{2}}+\frac{cn\pi^{\frac{n}{2}}M^{*n-a}}{\Gamma(\frac{n}{2}+1)}r}{\epsilon|\int_{\mathbb{R}^n}f(\boldsymbol{t})\,d\boldsymbol{t}|} \\
&+\frac{(M+M^*)^2\|h_i\|\left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\right)^{\frac{1}{2}}+\frac{cn\pi^{\frac{n}{2}}M^{*n+1-a}}{\Gamma(\frac{n}{2}+1)}u}{\epsilon|\int_{\mathbb{R}^n}f(\boldsymbol{t})\,d\boldsymbol{t}|}\}\,.
\end{aligned}
$$

In addition, for any $\sigma \geq \sigma^*$, and for any $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$, the following inequality holds.

$$
|\frac{\int_{\mathbb{R}^n}f(\boldsymbol{t})t_i\,d\boldsymbol{t}}{\int_{\mathbb{R}^n}f(\boldsymbol{t})\,d\boldsymbol{t}}-\frac{\int_{\mathbb{R}^n}f(\boldsymbol{t})t_ik(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}}{\int_{\mathbb{R}^n}f(\boldsymbol{t})k(\boldsymbol{x}-\boldsymbol{t};\sigma^2)\,d\boldsymbol{t}}| \leq \epsilon\,. \tag{2.211}
$$

Consequently, this result holds for all $i = 1, 2, \cdots n$ "simultaneously", when stated as the following. For $i = 1, 2, \cdots, n$, and any $\epsilon_i > 0$ there "always exists" some $\sigma^* > 0$ that satisfies the following inequality.

$$
\begin{aligned}
&\sigma^{*2} \geq \\
&\max\{\frac{(M+M^*)^2\|f\|\left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\right)^{\frac{1}{2}}+\frac{cn\pi^{\frac{n}{2}}M^{*n-a}}{\Gamma(\frac{n}{2}+1)}r}{|\int_{\mathbb{R}^n}f(\boldsymbol{t})\,d\boldsymbol{t}|} \\
&,\max_{i=1,\cdots,n}\{\frac{|\int_{\mathbb{R}^n}t_if(\boldsymbol{t})\,d\boldsymbol{t}|}{|\int_{\mathbb{R}^n}f(\boldsymbol{t})\,d\boldsymbol{t}|}\frac{(M+M^*)^2\|f\|\left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\right)^{\frac{1}{2}}+\frac{cn\pi^{\frac{n}{2}}M^{*n-a}}{\Gamma(\frac{n}{2}+1)}r}{\epsilon|\int_{\mathbb{R}^n}f(\boldsymbol{t})\,d\boldsymbol{t}|} \\
&+\frac{(M+M^*)^2\|h_i\|\left(\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}M^{*n}\right)^{\frac{1}{2}}+\frac{cn\pi^{\frac{n}{2}}M^{*n+1-a}}{\Gamma(\frac{n}{2}+1)}u}{\epsilon|\int_{\mathbb{R}^n}f(\boldsymbol{t})\,d\boldsymbol{t}|}\}\,\}\,.
\end{aligned}
$$

where $h_i(\boldsymbol{t}) \triangleq t_if(\boldsymbol{t})$. In addition, for any $\sigma \geq \sigma^*$, and for any $\boldsymbol{x} \in \mathcal{B}(\mathbf{0}, M)$ the following inequality holds.

$$\forall i \in \{1, 2, \cdots, n\} \; ; \; |\frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}| \leq \epsilon$$

$$\Rightarrow \quad \sum_{i=1}^{n} |\frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}| \leq n\epsilon_i \, . \tag{2.212}$$

On the other hand, we have the following fact,

$$\| \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}} \|_\infty \tag{2.213}$$

$$\leq \quad \sum_{i=1}^{n} |\frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}| \, . \tag{2.214}$$

From (2.212) and (2.213), the it follows that,

$$\| \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}} - \frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x} - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}} \|_\infty \leq n\epsilon \tag{2.215}$$

We stress that for any "arbitrarily small" $\epsilon > 0$, there always exists some corresponding σ^* that satisfies the bound in (2.215) for any $\sigma \geq \sigma^*$.

It just remains to show how (2.215) is related to a stationary point of $g(\boldsymbol{x}; \sigma)$. We proceed by writing down the definition of a stationary point $\boldsymbol{x}^*_\sigma$ as below.

$$\nabla g(\boldsymbol{x}^*_\sigma; \sigma) = \boldsymbol{0} \tag{2.216}$$

$$\equiv \quad \nabla [f(\,.\,) \star k(\,.\,; \sigma^2)](\boldsymbol{x}^*_\sigma) = \boldsymbol{0} \tag{2.217}$$

$$\equiv \quad [f(\,.\,) \star \nabla k(\,.\,; \sigma^2)](\boldsymbol{x}^*_\sigma) = \boldsymbol{0} \tag{2.218}$$

$$\equiv \quad \int_{\mathbb{R}^n} f(\boldsymbol{t}) \frac{(\boldsymbol{x}^*_\sigma - \boldsymbol{t})}{\sigma^2} k(\boldsymbol{x}^*_\sigma - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t} = \boldsymbol{0} \tag{2.219}$$

$$\equiv \quad \boldsymbol{x}^*_\sigma \int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x}^*_\sigma - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t} = \int_{\mathbb{R}^n} \boldsymbol{t} \, f(\boldsymbol{t}) k(\boldsymbol{x}^*_\sigma - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t} \tag{2.220}$$

$$\equiv \quad \boldsymbol{x}^*_\sigma = \frac{\int_{\mathbb{R}^n} \boldsymbol{t} \, f(\boldsymbol{t}) k(\boldsymbol{x}^*_\sigma - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) k(\boldsymbol{x}^*_\sigma - \boldsymbol{t}; \sigma^2) \, d\boldsymbol{t}} \, . \tag{2.221}$$

Plugging (2.221) into (2.215) proves the corollary.

$$\|\frac{\int_{\mathbb{R}^n} f(\boldsymbol{t}) t_i \, d\boldsymbol{t}}{\int_{\mathbb{R}^n} f(\boldsymbol{t}) \, d\boldsymbol{t}} - \boldsymbol{x}_\sigma^*\|_\infty \leq n\epsilon \tag{2.222}$$

□

CHAPTER 3

TRANSFORMATION KERNELS

It is one of the most fundamental problems in computer vision to establish alignment between images. This task is crucial for many important problems such as structure from motion, recognizing an object from different viewpoints, and tracking objects in videos. Roughly speaking, mainstream image alignment techniques can be categorized into "intensity-based" and "feature-based" methods. Intensity-based methods use dense pixel information (such as brightness pattern or correlation) integrated from image regions to estimate the geometric transformation [29]. In contrast, feature-based methods first extract a sparse set of local features from individual images, and then establish correspondence among them to infer the underlying transformation (for larger regions) [30].

In many applications intensity-based methods are appealing due to their direct access to richer information (i.e. to every single pixel) [29]. This can be useful, for example, when working with semi-regular patterns that are difficult to match by local features [31]. However, the practical performance of direct intensity methods can be undermined by the associated optimization challenge [32]. Specifically, it is well-known that optimizing a cost function that directly compares intensities of an image pair is highly susceptible to finding local minima [33]. Thus, unless very good initialization is provided, plain direct alignment of image intensities may lead to poor results.

Coarse-to-fine smoothing has become widely adopted to remedy the local minima issue in the alignment [34–39]. Despite its popularity, we will show that there are serious theoretical and practical issues with the Lucas-Kanade scheme when applied to non-translational motions.

In this work, we propose Gaussian smoothing the objective function of the alignment task, instead of the images, where the former was in fact the original goal of coarse-to-fine image smoothing techniques. In particular, we derive the theoretically correct image blur kernels that arise from (Gaussian)

smoothing an alignment objective function. We show that, for smoothing the objective of common motion models, such as affine and homography, there exists a corresponding ***integral operator in the image space*** . Thus, instead of convolving the objective function with a Gaussian kernel in transformation space, we can equivalently compute the integral transform in the image space. The former may be computationally expensive due to the ***curse of dimensionality*** for numerical integration. Our derived kernels are spatially varying as long as the transformation is not a pure translation, and vary from those heuristically suggested by [26] or [27].

We believe our rigorously derived results here could open up new investigations on the efficient and fast ways to approximate these kernels. Furthermore, the deterministic nature of the optimization algorithm provides a new opportunity for performance evaluation indices that are ***reproducible*** [1]. Most of the materials in this chapter are published in the paper [41].

3.1 Motivation

Lucas and Kanade made a major improvement to optimization of the direct intensity-based method by adopting a coarse-to-fine scheme [36]. The approach was motivated from a ***displacement*** alignment task as follows. Given a pair of images $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$. The goal is to estimate the optimal displacement $\boldsymbol{\theta}^*$ by solving $\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \int_{\mathcal{X}} (f_1(\boldsymbol{x}+\boldsymbol{\theta}) - f_2(\boldsymbol{x}))^2 d\boldsymbol{x}$. As mentioned earlier, this objective function may have a lot of local minima. However, if $f_1(\boldsymbol{x}+\boldsymbol{\theta})$ is linearized in $\boldsymbol{\theta}$ around the origin $\mathbf{0}$, the objective function becomes a ***convex*** quadratic with a closed form for the global minimum.

The quality of linearized approximate depends on the contribution of higher order terms. By Taylor's remainder theorem, higher order terms in the Taylor series of $f_1(\boldsymbol{x}+\boldsymbol{\theta})$ are negligible when either the "displacement" $\|\boldsymbol{\theta}\|$, or the induced norm of the Hessian $\|\nabla^2_{\boldsymbol{\theta}} f_1(\boldsymbol{x}+\boldsymbol{\theta})\|$ at any $\boldsymbol{\theta}$, is "very small". These two can be respectively achieved by reducing image resolution and reducing $\max_{\boldsymbol{x}\in\mathcal{X}} \|\nabla^2_{\boldsymbol{x}} f_1(\boldsymbol{x})\|$. The latter is typically achieved by

[1] In contrast, probabilistic schemes for robust fitting such as RANSAC [40] may produce a different answer in each run. The problem may persist even if a large batch of RANSAC solutions is aggregated and the best among them is selected.

isotropic Gaussian filtering. Observe that both of these transformations diminish the fine details in the image while leaving coarser structures intact, depending on the resolution or bandwidth parameter σ of the filter. The optimal displacement obtained in such coarse level is then used to warp f_1 accordingly. However, displacement correction obtained without seeing the details only provides a rough estimate of the actual displacement in the original images. Coarse-to-fine strategy iteratively alternates between correcting displacement in the lower resolution and moving to a finer resolution.

It was later shown that such a coarse-to-fine scheme is indeed guaranteed to recover the optimal ***displacement*** under some mild conditions [42]. Although a guarantee of correctness is only established for translational motion, the notion of coarse-to-fine smoothing followed by local approximation has been adopted in computer vision to matching with almost all parametric transformation models [34–39]. Despite its popularity, there are serious theoretical and practical issues with the Lucas-Kanade scheme when applied to non-translational motions. For example, if the transformation is scaling, it is easy to show that the Hessian of the image function may grow proportional to the distance from the origin. To compensate for this effect, stronger smoothing is required for points farther from the origin.

For example, consider scaling an image by a factor of θ. The linearization error is thus bounded by $|\partial^2/\partial\theta^2(f_1(\theta\boldsymbol{x}))| = |\boldsymbol{x}^T\nabla^2_{\boldsymbol{x}}f_1(\theta\boldsymbol{x})\boldsymbol{x}|$, which can can grow like $\|\boldsymbol{x}\|^2$. Thus, to keep $|\partial^2/\partial\theta^2(f_1(\theta\boldsymbol{x}))|$ small, say by some filtering process, the filter must shrink $\nabla^2_{\boldsymbol{x}}f_1(\theta\boldsymbol{x})$ more aggressively as traveling away from the origin. Obviously, this is not the case in Gaussian convolution where the amount of blur is constant everywhere.

We believe, the popularity of "***isotropic Gaussian convolution***" for image blurring is, in part, a legacy of ***scale-space theory***. [2]. This influential theory emerged from a series of seminal articles in the 80's [44–46]. This theory shows that "isotropic Gaussian convolution" is the "unique" linear operator obeying some least commitment axioms [47]. In particular, this operator is unbiased to location and orientation, due to its convolutional and isotropic nature. Later, Lindeberg extended scale-space theory to cover affine blur by anisotropic ***spatially invariant*** kernels [48, 49].

[2]The idea of Gaussian smoothing in vision is even older than scale-space theory. For example, Marr and Hildreth [43] studied zero-crossings the Laplacian in images convolved with Gaussian kernels at different scales.

Nevertheless, it is known that the human eye has progressively less resolution from the center (fovea) toward the periphery [50]. In computer vision, spatially varying blur is believed to benefit matching and alignment tasks. In that direction, Berg and Malik [26] introduced the notion of "geometric blur" and suggested some spatially varying kernels inspired by that. However, their kernels are derived heuristically, without a rigorous connection to the underlying geometric transformations. Some limitations of traditional image smoothing are discussed in [51] and coped using stacks of binary images. That work, however, still uses isotropic Gaussian kernel for smoothing images of the stack.

In this chapter, we propose Gaussian smoothing the objective function of the alignment task, instead of the images. In particular, we derive the theoretically correct image blur kernels that arise from (Gaussian) smoothing an alignment objective function. As we show, all of these kernels are spatially varying as long as the transformation is not a pure translation, and vary from those heuristically suggested by [26] or [27]. Besides theoretical benefits of deriving these kernels for deeper understanding of how blur and alignment are coupled, there is also a computational gain in using these kernels. Specifically, to achieve the smoothed objective, instead of convolving it with a Gaussian in transformation space, one can perform an integral transform in the image space. The latter has smaller dimensionality and thus cheaper to compute numerically.

3.2 Notation and Definitions

Let $\star$ and $\circledast$ denote convolution operators in spaces Θ and $\mathcal{X}$ respectively.

Given a signal $f : \mathcal{X} \to \mathbb{R}$, e.g. a 2D image, we define a signal warping or ***domain transformation*** parameterized by $\boldsymbol{\theta}$ as $\boldsymbol{\tau} : \mathcal{X} \times \Theta \to \mathcal{X}$. Here $\boldsymbol{\theta}$ is concatenation of all the parameters of a transformation. For example, in case of affine $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ with $\boldsymbol{x} \in \mathbb{R}^2$, $\boldsymbol{\theta}$ is a 6 dimensional vector containing the the elements of $\boldsymbol{A}$ and $\boldsymbol{b}$.

The ***Fourier transform*** of a real valued function $f : \mathbb{R}^n \to \mathbb{R}$ is $\hat{f}(\boldsymbol{\omega}) \triangleq \int_{\mathbb{R}^n} f(\boldsymbol{x}) e^{-i\boldsymbol{\omega}^T \boldsymbol{x}} d\boldsymbol{x}$ and the ***inverse Fourier transform*** is $\hat{f}(\boldsymbol{x}) = (2\pi)^{-n} \int_{\mathbb{R}^n} f(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}^T \boldsymbol{x}} d\boldsymbol{\omega}$.

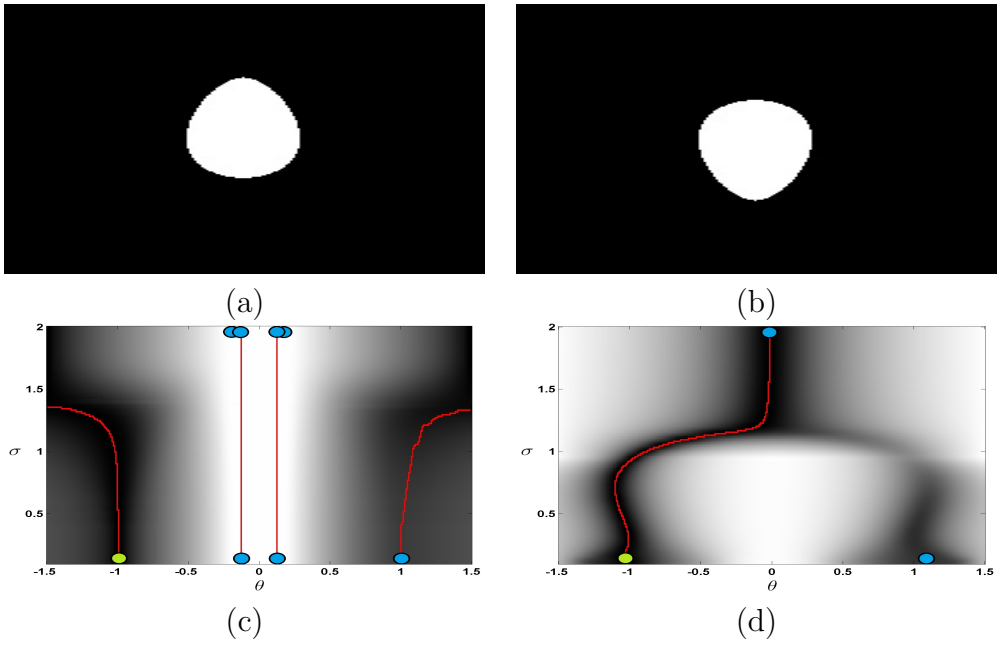


Figure 3.1: Basin of attraction for scale alignment. Egg shape input images are shown in (a) and (b), where black and white pixels are respectively by -1 and 1 intensity values. Obviously, the correct alignment is attained at $\theta = -1$, due to reflection symmetry. The objective function for z_{LK} is shown in (c) and for z in (d). Blue, green and red respectively indicate local maxima, global maximum and basin of attraction originating from local maxima of highest blur.

3.3 Smoothing the Objective

We use the inner product between the transformed f_1 and the reference signal f_2 as the alignment objective function. Note that f_1 and f_2 are the input to the alignment algorithm, and in many scenarios may be different from the original signals. For example, they may be mean subtracted or normalized by their ℓ_2 norm. The alignment objective function is denoted by $h(\boldsymbol{\theta})$ and defined as follows,

$$h(\boldsymbol{\theta}) \triangleq \int_{\mathcal{X}} f_1(\boldsymbol{\tau}(\boldsymbol{x}, \boldsymbol{\theta}))\, f_2(\boldsymbol{x})\, d\boldsymbol{x}\,, \tag{3.1}$$

where $f_1(\boldsymbol{\tau}(\boldsymbol{x}, \boldsymbol{\theta}))$ is signal f_1 warped by $\boldsymbol{\tau}(\boldsymbol{x}, \boldsymbol{\theta})$. Our goal is to find the parameters $\boldsymbol{\theta}^*$ that optimize the objective function (3.1). In practice, h may have multiple local optima. Thus, instead of directly optimizing h, we iteratively optimize a smoothed version of h in a coarse-to-fine approach.

Algorithm 2 Alignment by Gaussian Smoothing.

1: Input: $f_1 : \mathcal{X} \to \mathbb{R}$, $f_2 : \mathcal{X} \to \mathbb{R}$, $\boldsymbol{\theta}_0 \in \Theta$, ...
   The set $\{\sigma_k\}$ for $k = 1, \ldots, K$ s.t. $0 < \sigma_{k+1} < \sigma_k$
2: **for** $k = 1 \to K$ **do**
3: $\quad \boldsymbol{\theta}_k =$ local maximizer of $z(\boldsymbol{\theta}; \sigma_k)$ initialized at $\boldsymbol{\theta}_{k-1}$
4: **end for**
5: Output: $\boldsymbol{\theta}_K$

We denote the objective function $h(\boldsymbol{\theta})$ obtained after smoothing as $z(\boldsymbol{\theta}, \sigma)$, where σ determines the amount of smoothing. Given $z(\boldsymbol{\theta}, \sigma)$, we adopt the standard optimization approach described by Algorithm 2. That is, we use the parameters $\boldsymbol{\theta}_{k-1}$ found at a coarser scale to initialize the solution $\boldsymbol{\theta}_k$ found at each progressively finer scale.

In the Lucas-Kanade algorithm [36], instead of smoothing the objective function, they directly blur the images. This results in the following form for the objective function,

$$z_{LK}(\boldsymbol{\theta}, \sigma) \triangleq \int_{\mathcal{X}} [f_1(\boldsymbol{\tau}(\,.\,, \boldsymbol{\theta})) \circledast k(\,.\,; \sigma^2)]\, [f_2 \circledast k(\,.\,; \sigma^2)](\boldsymbol{x})\; d\boldsymbol{x}\,.$$

Image smoothing is done in hope of eliminating the brittle local optima in the objective function. However, if the latter is our goal, we propose the correct approach is to blur the objective function directly[3],

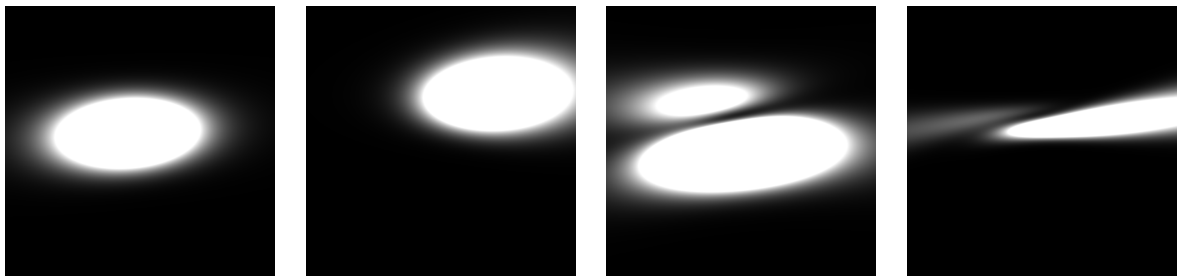
$$z(\boldsymbol{\theta}, \sigma) \triangleq [h \star k(\,\cdot\,, \sigma^2)](\boldsymbol{\theta})\,.$$

The optimization landscape of these two cases may differ significantly. To illustrate, consider the egg shape images in Figures 3.1(a) and 3.1(b). If we assume the only parameter subject to optimization is the scale factor, i.e. $\boldsymbol{\tau}(\boldsymbol{x}, \theta) = \theta \boldsymbol{x}$, the associated optimization landscape is visualized in figure 3.1(c) for z_{LK} and 3.1(d) for z. Clearly, z has a single basin of attraction that leads to the global optimum, unlike z_{LK} whose basins do not necessarily land at the global optimum.

[3] $[h \star k(\,\cdot\,, \sigma^2)](\boldsymbol{\theta})$ is bounded when either signals decay rapidly enough or have bounded support. In image scenario, the latter always holds.

Name	$\boldsymbol{\theta}$	$\boldsymbol{\tau}(\boldsymbol{x}, \boldsymbol{\theta})$	$u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})$
Translation	$\boldsymbol{d}_{n\times 1}$	$\boldsymbol{x} + \boldsymbol{d}$	$k(\boldsymbol{\tau}(\boldsymbol{x}, \boldsymbol{\theta}) - \boldsymbol{y}; \sigma^2)$
Translation+Scale	$[\, \boldsymbol{a}_{n\times 1}\,,\, \boldsymbol{d}_{n\times 1}\,]$	$\boldsymbol{a}^T\boldsymbol{x} + \boldsymbol{d}$	$K(\boldsymbol{\tau}(\boldsymbol{x}, \boldsymbol{\theta}) - \boldsymbol{y}; \sigma^2\,\mathrm{diag}([1 + x_i^2]))$
Affine	$[\, vec(\boldsymbol{A}_{n\times n})\,,\, \boldsymbol{b}_{n\times 1}\,]$	$\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$	$k(\boldsymbol{\tau}(\boldsymbol{x}, \boldsymbol{\theta}) - \boldsymbol{y}; \sigma^2(1 + \|\boldsymbol{x}\|^2))$
Homography	$[\, vec(\boldsymbol{A}_{n\times n})\,,\, \boldsymbol{b}_{n\times 1}\,,\, \boldsymbol{c}_{n\times 1}\,]$	$\frac{1}{1+\boldsymbol{c}^T\boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$	$q(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y}, \sigma)\; e^{-p(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y}, \sigma)}$

Table 3.1: Kernels for some of the common transformations arising in vision (for all kernels $n \geq 1$ except homography where $n = 2$).



(a) (b) (c) (d)

Figure 3.2: Visualization of affine and homography kernels specified by $\boldsymbol{A}_0 = [2 \quad 0.2\,;\, -0.3 \quad 4], \boldsymbol{b}_0 = [0.15 \quad -0.25]$ (also $\boldsymbol{c}_0 = [1 \quad -5]$ for homography). Here $\boldsymbol{x} \in [-1, 1] \times [-1, 1]$, $\sigma = 0.5$ and $\boldsymbol{y} = (1, 1)$ or $\boldsymbol{y} = (0, 0)$. More precisely, affine kernels in (a) $u(\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{x}, \boldsymbol{y} = (0, 0))$ (b) $u(\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{x}, \boldsymbol{y} = (1, 1))$ and homography kernels in (c) $u(\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{x}, \boldsymbol{y} = (0, 0))$ (d) $u(\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{x}, \boldsymbol{y} = (1, 1))$.

3.4 Transformation Kernels

Our goal is to perform optimization on the ***smoothed*** objective function. Smoothing the objective function refers to a convolution in the space of transformation parameters with a Gaussian kernel. Unfortunately, performing this convolution may be computationally expensive when the dimensionality of the ***transformation space*** is large, e.g. eight for homography of 2D images. This section introduces the notion of transformation kernels, which enables us to equivalently write the smoothed objective function using some ***integral transform*** of the signal. This integration is performed in the ***image space*** (e.g. 2D for images), reducing the computational complexity.

Definition Given a domain transformation $\boldsymbol{\tau} : \mathcal{X} \times \Theta \to \mathcal{X}$, where $\mathcal{X} = \mathbb{R}^n$ and $\Theta = \mathbb{R}^m$. We define the ***transformation kernel*** associated with $\boldsymbol{\tau}$ as $u_{\boldsymbol{\tau},\sigma} : \Theta \times \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ to be the function satisfying the following ***integral***

equation for ***all*** Schwartz[4] functions f,

$$[f(\boldsymbol{\tau}(\boldsymbol{x},\cdot)) \star k(\cdot;\sigma^2)]\,(\boldsymbol{\theta}) = \int_{\mathcal{X}} f(\boldsymbol{y})\, u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y}) d\boldsymbol{y}\,. \tag{3.2}$$

Using this definition, the smoothed alignment objective z can be equivalently written as the following,

$$\begin{aligned}
& z(\boldsymbol{\theta},\sigma) && (3.3)\\
\triangleq\ & [h \star k(\,\cdot\,,\sigma^2)](\boldsymbol{\theta}) && (3.4)\\
=\ & \int_{\mathcal{X}} \big(f_2(\boldsymbol{x})[f_1(\boldsymbol{\tau}(\boldsymbol{x},.)) \star k(\,\cdot\,,\sigma^2)](\boldsymbol{\theta})\,\big)\, d\boldsymbol{x} && (3.5)\\
=\ & \int_{\mathcal{X}} \left(f_2(\boldsymbol{x}) \left(\int_{\mathcal{X}} f_1(\boldsymbol{y}) u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y})\, d\boldsymbol{y}\right)\right) d\boldsymbol{x}\,, && (3.6)
\end{aligned}$$

where the integral transform in (3.6) uses the definition of kernel provided in (3.2). A procedure for computing the integral transform (3.6) will be provided in section 3.5.

3.4.1 Derivation of Kernels

Proposition 14 *The following choice of u is a solution to the definition of a kernel provided in (3.2). Here $\mathcal{X} = \Omega = \mathbb{R}^n$.*

$$\begin{aligned}
& u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y})\\
=\ & \frac{1}{(2\pi)^n} \int_{\Omega} \Big(\int_{\Theta} e^{i\boldsymbol{\omega}^T(\boldsymbol{\tau}(\boldsymbol{x},\boldsymbol{t})-\boldsymbol{y})} k(\boldsymbol{t}-\boldsymbol{\theta};\sigma^2)\, d\boldsymbol{t}\Big) d\boldsymbol{\omega} && (3.7)
\end{aligned}$$

The proof uses the Fourier representation $f(\boldsymbol{x}) = (2\pi)^{-n} \int_{\Omega} \hat{f}(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}^T\boldsymbol{x}}\, d\boldsymbol{\omega}$, and then application of ***Parseval's theorem*** . See the appendix for details.

Now by applying the result of proposition 14 to the desired transformation $\boldsymbol{\tau}$, we can compute the integrals[5] and derive the corresponding kernel function as shown in Table 3.1 (see also figure 3.2 for some visualization). The

[4] A Schwartz function is one whose derivatives are rapidly decreasing.

[5] Although the integral in (3.7) does not necessarily have a "closed-form" for any arbitrary transformation $\boldsymbol{\tau}$, it does so for most of the transformations we care about in practice, as listed in Table 3.1.

functions q and p, associated with the homography kernel, are each a ratio of polynomials[6].

The complete derivation of these kernels is provided in the appendix. Nevertheless, below we present a relatively easy way to check the correctness of the kernels. Specifically, we check two necessary conditions of the heat equation and the limit behavior, which must hold for the kernels.

Heat Equation

Consider the convolution $[f(\boldsymbol{\tau}(\boldsymbol{x}, \cdot)) \star k(\cdot\,; \sigma)](\boldsymbol{\theta})$. Such Gaussian convolution obeys the ***heat equation*** [28]:

$$\begin{aligned} & \sigma \Delta_{\boldsymbol{\theta}} [f(\boldsymbol{\tau}(\boldsymbol{x}, \cdot)) \star k(\cdot\,; \sigma)](\boldsymbol{\theta}) \\ = \;& (\partial / \partial \sigma) [f(\boldsymbol{\tau}(\boldsymbol{x}, \cdot)) \star k(\cdot\,; \sigma)](\boldsymbol{\theta})\,. \end{aligned} \tag{3.8}$$

Since we argue that $[f(\boldsymbol{\tau}(\boldsymbol{x}, \cdot)) \star k(\cdot\,; \sigma)](\boldsymbol{\theta}) = \int_{\mathcal{X}} f(\boldsymbol{y}) u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y}$, the following must hold:

$$\begin{aligned} & \sigma \Delta_{\boldsymbol{\theta}} \int_{\mathcal{X}} f(\boldsymbol{y}) u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\ = \;& \frac{\partial}{\partial \sigma} \int_{\mathcal{X}} f(\boldsymbol{y}) u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} && (3.9) \\ \equiv \;& \int_{\mathcal{X}} f(\boldsymbol{y}) \sigma \Delta_{\boldsymbol{\theta}} u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\ = \;& \int_{\mathcal{X}} f(\boldsymbol{y}) \frac{\partial}{\partial \sigma} u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} && (3.10) \\ \Leftarrow \;& \sigma \Delta_{\boldsymbol{\theta}} u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y}) = \frac{\partial}{\partial \sigma} u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})\,, && (3.11) \end{aligned}$$

[6]Complete expression for the homography kernel qe^{-p} is as below:

$$\begin{aligned} \gamma_0 &\triangleq \frac{1}{1 + \|\boldsymbol{x}\|^2} \\ \gamma_1 &\triangleq 1 + \boldsymbol{c}^T \boldsymbol{x} \\ \boldsymbol{v} &\triangleq \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} \\ q &\triangleq \gamma_0 \frac{(\gamma_0 \|\boldsymbol{x}\|^2 \boldsymbol{y}^T \boldsymbol{v} + \gamma_1)^2 + \sigma^2 \|\boldsymbol{x}\|^2 (1 + \gamma_0 \|\boldsymbol{x}\|^2 \|\boldsymbol{y}\|^2)}{2\pi\sigma^2 (1 + \gamma_0 \|\boldsymbol{x}\|^2 \|\boldsymbol{y}\|^2)^{\frac{5}{2}}} \\ p &\triangleq \frac{\|\gamma_1 \boldsymbol{y} - \boldsymbol{v}\|^2 + \gamma_0 \|\boldsymbol{x}\|^2 (v_2 y_1 - v_1 y_2)^2}{2\sigma^2 (1 + \|\boldsymbol{x}\|^2 (1 + \|\boldsymbol{y}\|^2))}\,. \end{aligned}$$

where $\Leftarrow$ in (3.11) means sufficient condition. Now it is much easier to check the identity (3.11) for the provided kernels. For example, in the case of an affine kernel $k(\boldsymbol{\tau}(\boldsymbol{x},\boldsymbol{\theta})-\boldsymbol{y};\sigma^2(1+\|\boldsymbol{x}\|^2))$, both sides of the identity are equal to $(\frac{\|\boldsymbol{\tau}(\boldsymbol{x},\boldsymbol{\theta})-\boldsymbol{y}\|^2}{\sigma^3(1+\|\boldsymbol{x}\|^2)}-\frac{n}{\sigma})\ k(\boldsymbol{\tau}(\boldsymbol{x},\boldsymbol{\theta})-\boldsymbol{y};\sigma^2(1+\|\boldsymbol{x}\|^2))$.

Limit Behavior

When the amount of smoothing approaches zero, the integral transform must recover the original function. Formally, we want the following identity to hold,

$$\lim_{\sigma\to 0^+}\int_{\mathcal{X}} f(\boldsymbol{y})u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y} = f(\boldsymbol{\tau}(\boldsymbol{x},\boldsymbol{\theta}))\,. \tag{3.12}$$

The sufficient condition for the above identity is that $\lim_{\sigma\to 0^+}\ u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y}) = \delta(\boldsymbol{\tau}(\boldsymbol{x},\boldsymbol{\theta})-\boldsymbol{y})$, where δ is Dirac's delta function. This is trivial for the kernels of affine and its special cases; since the kernel itself is a Gaussian, $\lim_{\sigma\to 0^+}$ is equivalent to kernel's variance approaching to zero (for any bounded choice of $\|\boldsymbol{x}\|$). It is known that when the variance of the normal density function tends to zero it approaches Dirac's delta function.

3.4.2 Remarks

Two interesting observations can be made about Table 3.1. First, from a ***purely objective*** standpoint, the derived kernels exhibit "foveation", similar to that in the eye. Except for translation, all the kernels are spatially varying with density decreasing in $\|\boldsymbol{x}\|$. This is very easy to check for translation+scale and affine kernels, where they are spatially varying Gaussian kernels whose variance depends and increases in $\|\boldsymbol{x}\|$.

Second observation is about the geometric blur kernel proposed by Berg and Malik [26], which has the form $u_\sigma(\boldsymbol{x},\boldsymbol{y}) = k(\boldsymbol{y}-\boldsymbol{x};\sigma^2\|\boldsymbol{x}\|^2)$. The Berg and Malik's kernel is only a restricted case of our kernels; when the transformation is restricted to the identity transform $\boldsymbol{\tau}(\boldsymbol{x}) = \boldsymbol{x}$. Another deficiency of Berg and Malik's kernel is that it becomes singular as $\|\boldsymbol{x}\| \to 0$, while the proposed kernel remain stable. Finally, it is not clear how Berg and Malik's kernel affects the optimization landscape of the matching or alignment task. However, it is transparent that our proposed kernels listed in Table 3.1 smooth the optimization landscape in Gaussian sense. In fact, to the best of

our knowledge, this work is the first that rigorously derives kernels for such transformations.

3.4.3 Image Blurring vs. Objective Blurring

It is now easy to check that for the "translation transformation", Gaussian convolution of the alignment objective with respect to the optimization variables is equivalent to applying a "Gaussian convolution" to the image f_1. This is easy to check by plugging the translation kernel from Table 3.1 into the smoothed objective function (3.6) as below:

$$
\begin{aligned}
& z(\boldsymbol{\theta}, \sigma) && (3.13)\\
&= \int_{\mathcal{X}} \left(f_2(\boldsymbol{x}) \int_{\mathcal{X}} f_1(\boldsymbol{y}) u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \right) d\boldsymbol{x} && (3.14)\\
&= \int_{\mathcal{X}} \left(f_2(\boldsymbol{x}) \int_{\mathcal{X}} f_1(\boldsymbol{y}) k(\boldsymbol{\theta} + \boldsymbol{x} - \boldsymbol{y}; \sigma^2)\, d\boldsymbol{y} \right) d\boldsymbol{x} && (3.15)\\
&= \int_{\mathcal{X}} \left(f_2(\boldsymbol{x}) \left[f_1(\,\cdot\,) \circledast k(\,\cdot\,; \sigma^2) \right] (\boldsymbol{\theta} + \boldsymbol{x}) \right) d\boldsymbol{x}\,. && (3.16)
\end{aligned}
$$

However, such equivalence does not hold for other transformations, e.g. affine. There, Gaussian convolution of the alignment objective with respect to the optimization variables is equivalent to an "integral transform" of f_1, which cannot be expressed by the convolution of f_1 with some spatially invariant convolution kernel in image space as shown below for affine case:

$$
\begin{aligned}
& z(\boldsymbol{\theta}, \sigma) && (3.17)\\
&= \int_{\mathcal{X}} \left(f_2(\boldsymbol{x}) \int_{\mathcal{X}} f_1(\boldsymbol{y}) u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \right) d\boldsymbol{x} && (3.18)\\
&= \int_{\mathcal{X}} \int_{\mathcal{X}} \frac{f_2(\boldsymbol{x}) f_1(\boldsymbol{y}) e^{-\frac{\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}-\boldsymbol{y}\|^2}{2\sigma^2(1+\|\boldsymbol{x}\|^2)}}}{(\sigma^2 2\pi(1+\|\boldsymbol{x}\|^2))^{\frac{n}{2}}} d\boldsymbol{y} d\boldsymbol{x}. && (3.19)
\end{aligned}
$$

3.5 Computation of the Integral Transform

Kernels can offer computational efficiency when computing the smoothed objective (3.3).

If the kernel u is affine or one of its special cases, then it is a Gaussian form[7] in variable $\boldsymbol{y}$ according to Table 3.1. In such cases, expressing f_1 by Gaussian Basis Functions[8], piecewise constant or piecewise polynomial forms leads to a closed form of the integral transform. Details are provided in sections 3.5.1 and 3.5.2.

If the kernel u is not Gaussian in $\boldsymbol{y}$ (such as in homography), the derivation of a closed form for the integral transform may not be possible. However, numerical integration is done much more efficiently using the kernelized form (3.6) compared to the original form (3.3). For example, when $n = 2$, integration in the original form is over $\boldsymbol{\theta}$ and for homography $\dim(\boldsymbol{\theta}) = 8$. However, the equivalent integral transform is over $\boldsymbol{y}$, where $\dim(\boldsymbol{y}) = 2$.

3.5.1 Gaussian RBF Representation of f_1

The following result addresses the representation of f_1 by Gaussian Radial Basis Functions (GRBFs) $\phi(\boldsymbol{x}; \boldsymbol{x}_0, \delta_0) = e^{-\frac{\|\boldsymbol{x}-\boldsymbol{x}_0\|^2}{2\delta_0^2}}$; the more general case of GBFs can be obtained in a similar fashion.

Proposition 15 *Suppose* $f_1 = \sum_{k=1}^{p} a_k \phi(\boldsymbol{y}; \boldsymbol{x}_k, \delta_k)$, *where* $\phi(\boldsymbol{x}; \boldsymbol{x}_k, \delta_k) = e^{-\frac{\|\boldsymbol{x}-\boldsymbol{x}_k\|^2}{2\delta_k^2}}$. *Assume that* $u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})$ *is Gaussian in variable* $\boldsymbol{y}$. *Then the following identity holds.*

$$\int_{\mathcal{X}} f_1(\boldsymbol{y}) u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} = \sum_{i=1}^{p} a_i \left(\frac{\delta_i}{\sqrt{\delta_i^2 + s^2}}\right)^n e^{-\frac{\|\boldsymbol{x}_i-\boldsymbol{\tau}\|^2}{2(\delta_i^2+s^2)}}.$$

See the appendix for a proof.

3.5.2 Piecewise Constant Representation of f_1

The following result addresses the representation of f_1 as piecewise constant; the extension to piecewise polynomial is straightforward.

[7] We say a kernel is Gaussian in $\boldsymbol{y}$ when it can be written as $u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y}) = k(\boldsymbol{\tau}(\boldsymbol{\theta}, \boldsymbol{x}) - \boldsymbol{y}; s^2(\boldsymbol{\theta}, \boldsymbol{x}))$, where $s : \Theta \times \mathcal{X} \to \mathbb{R}_+$ is an arbitrary map and the maps $\boldsymbol{\tau}$ and s are ***independent*** of $\boldsymbol{y}$.

[8] A GBF is a function of form $\Phi(\boldsymbol{x}; \boldsymbol{x}_0, \boldsymbol{\Delta}_0) = \exp(-\frac{(\boldsymbol{x}-\boldsymbol{x}_0)^T \boldsymbol{\Delta}_0^{-1} (\boldsymbol{x}-\boldsymbol{x}_0)}{2})$, where the matrix $\boldsymbol{\Delta}$ is positive definite. It is known that Gaussian RBFs $\phi(\boldsymbol{x}; \boldsymbol{x}_0, \delta_0) = \exp(-\frac{\|\boldsymbol{x}-\boldsymbol{x}_0\|^2}{2\delta_0^2})$, which are a special case of GBFs, are ***general function approximators*** .

Proposition 16 *Suppose* $f_1(\boldsymbol{x}) = c$ *on a rectangular piece* $\boldsymbol{x} \in \mathcal{X}^\dagger \triangleq \Pi_{k=1}^n [\underline{x_k}, \overline{x_k}]$. *Assume that* $u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})$ *has the form* $K(\boldsymbol{q_\tau}(\boldsymbol{\theta}) - \boldsymbol{y}; \boldsymbol{S})$, *where* $\boldsymbol{S} \triangleq \mathrm{diag}(s_1^2, \cdots, s_n^2)$ *and* $\boldsymbol{q_\tau} : \Theta \to \mathbb{R}^n$ *is some map. Then the following identity holds:*

$$\begin{aligned} & \int_{\mathcal{X}^\dagger} f_1(\boldsymbol{y}) u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\ = & \prod_{k=1}^{n} \frac{1}{2} \Big(\mathrm{erf}\,\big(\frac{q_{\boldsymbol{\tau}\,k} - \overline{x_k}}{\sqrt{2} s_k}\big) - \mathrm{erf}\,\big(\frac{q_{\boldsymbol{\tau}\,k} - \underline{x_k}}{\sqrt{2} s_k}\big) \Big). \end{aligned}$$

The proof uses separability of integrals for diagonal K.

3.6 Regularization

Regularization may compensate for the numerical instability caused by excessive smoothing of the objective function and ***improve the well-posedness*** of the task. The latter means if there are multiple transformations that lead to equally good alignments (e.g. when image content has ***symmetries***), the regularization prefers the closest transformation to some given $\boldsymbol{\theta}_0$. This makes existence of a ***unique global optimum*** more presumable. We achieve these goals by replacing f_1 with the following regularized version:

$$\tilde{f}_1(\boldsymbol{\tau}(\cdot\,,\cdot\,), \boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\theta}_0, r) \triangleq k(\boldsymbol{\theta} - \boldsymbol{\theta}_0; r^2) f_1(\boldsymbol{\tau}(\boldsymbol{x}; \boldsymbol{\theta})). \tag{3.20}$$

Regularization shrinks the signal f_1 for peculiar transformations with very large $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|$. Typically $\boldsymbol{\theta}_0$ is set to the identity transformation $\boldsymbol{\tau}(\boldsymbol{x}; \boldsymbol{\theta}_0) = \boldsymbol{x}$. Using (3.20), the regularized objective function can be written as below:

$$\begin{aligned} \tilde{h}(\boldsymbol{\theta}; \boldsymbol{\theta}_0, r) \; & \triangleq \int_{\mathcal{X}} \Big(\tilde{f}_1(\boldsymbol{\tau}, \boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\theta}_0, r) f_2(\boldsymbol{x}) \Big)\, d\boldsymbol{x} \\ & = \int_{\mathcal{X}} k(\boldsymbol{\theta} - \boldsymbol{\theta}_0; r^2) f_1(\boldsymbol{\tau}(\boldsymbol{x}; \boldsymbol{\theta})) f_2(\boldsymbol{x})\, d\boldsymbol{x}\,. \end{aligned}$$

Consequently, the smoothed ***regularized*** objective is as follows,

$$\tilde{z}(\boldsymbol{\theta}, \boldsymbol{\theta}_0, r, \sigma) \triangleq [\tilde{h}(\cdot\,, \boldsymbol{\theta}_0, r) \star k(\cdot\,; \sigma^2)](\boldsymbol{\theta}). \tag{3.21}$$

This form is still amenable to kernel computation using the following proposition.

Proposition 17 *The regularized objective function $\tilde{z}(\boldsymbol{\theta}, \boldsymbol{\theta}_0, r, \sigma)$ can be written using transformation kernels as follows.*

$$
\begin{aligned}
& \tilde{z}(\boldsymbol{\theta}, \boldsymbol{\theta}_0, r, \sigma) && (3.22)\\
= \; & [\tilde{h}(\,\cdot\,, \boldsymbol{\theta}_0, r) \star k(\,\cdot\,; \sigma^2)](\boldsymbol{\theta}) && (3.23)\\
= \; & \int_{\mathcal{X}} \Big(k(\boldsymbol{\theta} - \boldsymbol{\theta}_0; r^2 + \sigma^2) f_2(\boldsymbol{x}) ... && (3.24)\\
& \cdot \int_{\mathcal{X}} \Big(f_1(\boldsymbol{y}) u_{\boldsymbol{\tau}, \frac{r\sigma}{\sqrt{r^2+\sigma^2}}} (\frac{r^2\boldsymbol{\theta} + \sigma^2\boldsymbol{\theta}_0}{r^2 + \sigma^2}, \boldsymbol{x}, \boldsymbol{y}) \Big) \, d\boldsymbol{y} \Big) d\boldsymbol{x} \, .
\end{aligned}
$$

See the appendix for the proof.

3.7 Proofs

3.7.1 Notation

The symbol $\triangleq$ is used for equality by definition. Also, we use x for scalars, $\boldsymbol{x}$ for vectors, $\boldsymbol{X}$ for matrices, and $\mathcal{X}$ for sets. In addition, $f(.)$ denotes a scalar valued function and $\boldsymbol{f}(.)$ a vector valued function. Unless stated otherwise, $\|\boldsymbol{x}\|$ means $\|\boldsymbol{x}\|_2$ and ∇ means $\nabla_{\boldsymbol{x}}$. Finally, $\star$ and $\circledast$ denote convolution operators in spaces Θ and $\mathcal{X}$ respectively.

3.7.2 Definitions

Definition [Domain Transformation] Given a function $f : \mathcal{X} \to \mathbb{R}$ and a vector field $\boldsymbol{\tau} : \mathcal{X} \times \Theta \to \mathcal{X}$, where $\mathcal{X} = \mathbb{R}^n$ and $\Theta = \mathbb{R}^m$. We refer to $\boldsymbol{\tau}(\boldsymbol{x}, \boldsymbol{\theta})$ as the ***domain transformation*** parameterized by $\boldsymbol{\theta}$. Note that the parameter vector $\boldsymbol{\theta}$ is constructed by concatenation of all the parameters of a transformation. For example, in case of affine $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ with $\boldsymbol{x} \in \mathbb{R}^2$, $\boldsymbol{\theta}$ is a 6 dimensional vectors containing the elements of $\boldsymbol{A}$ and $\boldsymbol{b}$.

Definition [Isotropic Gaussian]

$$k(\boldsymbol{x}; \sigma^2) \triangleq \frac{1}{(\sqrt{2\pi}\sigma)^{\dim(\boldsymbol{x})}} e^{-\frac{\|\boldsymbol{x}\|^2}{2\sigma^2}} . \tag{3.25}$$

Definition [Anisotropic Gaussian]

$$K(\boldsymbol{x}; \boldsymbol{\Sigma}) \triangleq \frac{1}{(\sqrt{2\pi})^{\dim(\boldsymbol{x})} \sqrt{\det(\boldsymbol{\Sigma})}} e^{-\frac{\boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x}}{2}} .$$

Definition [Fourier Transform]

We use the following convention for Fourier transform. The Fourier transform of a real valued function $f : \mathbb{R}^n \to \mathbb{R}$ is $\hat{f}(\boldsymbol{\omega}) = \int_{\mathbb{R}^n} f(\boldsymbol{x}) e^{-i\boldsymbol{\omega}^T \boldsymbol{x}} d\boldsymbol{x}$ and the inverse Fourier transform is $\hat{f}(\boldsymbol{x}) = (2\pi)^{-n} \int_{\mathbb{R}^n} f(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}^T \boldsymbol{x}} d\boldsymbol{\omega}$.

Definition [Transformation Kernel]

Given a domain transformation $\boldsymbol{\tau} : \Theta \times \mathcal{X} \times \Theta \to \mathcal{X}$, where $\mathcal{X} = \mathbb{R}^n$ and $\Theta = \mathbb{R}^m$. We define a ***transformation kernel*** associated with $\boldsymbol{\tau}$ as $u_{\boldsymbol{\tau},\sigma} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that it satisfies the following ***integral equation*** ,

$$\begin{aligned} &\forall f \quad : \\ &\left[f(\boldsymbol{\tau}(\boldsymbol{x}, \cdot)) \star k(\cdot; \sigma^2) \right] (\boldsymbol{\theta}) = \int_{\mathcal{X}} f(\boldsymbol{y})\, u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y}) d\boldsymbol{y} , \end{aligned} \tag{3.26}$$

where f is assumed to be a Schwartz function. Therefore, any transformation kernel that satisfies this equation allows the convolution of the transformed signal with the Gaussian kernel be equivalently written by the ***integral transform*** of the non-transformed signal with the kernel $u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})$.

Definition [Smoothed Regularized Objective]

We define the smoothed ***regularized*** objective as the following.

$$\tilde{z}(\boldsymbol{\theta}, \boldsymbol{\theta}_0, r, \sigma) \triangleq [\tilde{h}(\,\cdot\,, \boldsymbol{\theta}_0, r) \star k(\,\cdot\,; \sigma^2)](\boldsymbol{\theta}). \tag{3.27}$$

3.7.3 Proofs

Proposition 0 *The following identity holds for the product of two Gaussians.*

$$k(\boldsymbol{\tau}-\boldsymbol{\mu}_1; \sigma_1^2)\; k(\boldsymbol{\tau}-\boldsymbol{\mu}_2; \sigma_2^2) = \frac{e^{-\frac{\|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2\|^2}{2(\sigma_1^2+\sigma_2^2)}}}{(\sqrt{2\pi(\sigma_1^2+\sigma_2^2)})^m} k(\boldsymbol{\tau}-\frac{\sigma_2^2\boldsymbol{\mu}_1+\sigma_1^2\boldsymbol{\mu}_2}{\sigma_1^2+\sigma_2^2}; \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2})\,.$$

Proof Sketch

$$\begin{aligned}
& k(\boldsymbol{\tau}-\boldsymbol{\mu}_1; \sigma_1^2)\; k(\boldsymbol{\tau}-\boldsymbol{\mu}_2; \sigma_2^2) \\
= \;& \frac{1}{(\sigma_1\sqrt{2\pi})^m} e^{-\frac{\|\boldsymbol{\tau}-\boldsymbol{\mu}_1\|^2}{2\sigma_1^2}}\; \frac{1}{(\sigma_2\sqrt{2\pi})^m} e^{-\frac{\|\boldsymbol{\tau}-\boldsymbol{\mu}_2\|^2}{2\sigma_2^2}} \\
= \;& \frac{1}{(2\pi\sigma_1\sigma_2)^m} e^{-\frac{\|\boldsymbol{\tau}-\boldsymbol{\mu}_1\|^2}{2\sigma_1^2} - \frac{\|\boldsymbol{\tau}-\boldsymbol{\mu}_2\|^2}{2\sigma_2^2}} \\
= \;& \frac{1}{(2\pi\sigma_1\sigma_2)^m} e^{-\frac{\|\boldsymbol{\tau}-\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}(\frac{\boldsymbol{\mu}_1}{\sigma_1^2}+\frac{\boldsymbol{\mu}_2}{\sigma_2^2})\|^2}{2\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} - \frac{\|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2\|^2}{2(\sigma_1^2+\sigma_2^2)}} \\
= \;& \frac{e^{-\frac{\|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2\|^2}{2(\sigma_1^2+\sigma_2^2)}}}{(\sqrt{2\pi(\sigma_1^2+\sigma_2^2)})^m} k(\boldsymbol{\tau}-\frac{\sigma_2^2\boldsymbol{\mu}_1+\sigma_1^2\boldsymbol{\mu}_2}{\sigma_1^2+\sigma_2^2}; \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2})\,.
\end{aligned} \tag{3.28}$$

Note that (3.28) is derived by completing the square.

□

Proposition 18 *The following choice of u,*

$$\begin{aligned} & u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y}) \\ = \; & \frac{1}{(2\pi)^n}\int_{\Omega}\Big(\int_{\Theta} e^{i\boldsymbol{\omega}^T(\boldsymbol{\tau}(\boldsymbol{x},\boldsymbol{t})-\boldsymbol{y})}k(\boldsymbol{t}-\boldsymbol{\theta};\sigma^2)\,d\boldsymbol{t}\Big)d\boldsymbol{\omega} \end{aligned} \tag{3.29}$$

is a solution to the definition of kernel provided in (3.26). Here $\mathcal{X}=\Omega=\mathbb{R}^n$, and $k(\boldsymbol{t};\sigma^2)$ is some function $k(\,.\,;\sigma):\mathcal{X}\to\mathbb{R}$ with some parameter σ, which in our case is simply an isotropic Gaussian with bandwidth σ.

Proof Sketch The key to the proof is writing $f(\boldsymbol{x})$ by its Fourier form $f(\boldsymbol{x})=(2\pi)^{-n}\int_{\Omega}\hat{f}(\boldsymbol{\omega})e^{i\boldsymbol{\omega}^T\boldsymbol{x}}\,d\boldsymbol{\omega}$, where $\Omega=\mathbb{R}^n$ (similar to $\mathcal{X}=\mathbb{R}^n$).

$$\begin{aligned} & [f(\boldsymbol{\tau}(\boldsymbol{x},.))\overset{\Theta}{\circledast}k(\,.\,\sigma^2)](\boldsymbol{\theta}) \\ = \; & [\left(\frac{1}{(2\pi)^n}\int_{\Omega}\hat{f}(\boldsymbol{\omega})e^{i\boldsymbol{\omega}^T\boldsymbol{\tau}(\boldsymbol{x},.)}\,d\boldsymbol{\omega}\right)\overset{\Theta}{\circledast}k(\,.\,\sigma^2)](\boldsymbol{\theta}) \\ = \; & \frac{1}{(2\pi)^n}\int_{\Theta}\left(\int_{\Omega}\hat{f}(\boldsymbol{\omega})e^{i\boldsymbol{\omega}^T\boldsymbol{\tau}(\boldsymbol{x},\boldsymbol{t})}\,d\boldsymbol{\omega}\right)k(\boldsymbol{t}-\boldsymbol{\theta};\sigma^2)\,d\boldsymbol{t} \\ = \; & \frac{1}{(2\pi)^n}\int_{\Omega}\hat{f}(\boldsymbol{\omega})\left(\int_{\Theta}e^{i\boldsymbol{\omega}^T\boldsymbol{\tau}(\boldsymbol{x},\boldsymbol{t})}k(\boldsymbol{t}-\boldsymbol{\theta};\sigma^2)\,d\boldsymbol{t}\right)d\boldsymbol{\omega} \\ = \; & \frac{1}{(2\pi)^n}\int_{\mathcal{X}}f(\boldsymbol{y})\left(\int_{\Omega}e^{-i\boldsymbol{\omega}^T\boldsymbol{y}}\left(\int_{\Theta}e^{i\boldsymbol{\omega}^T\boldsymbol{\tau}(\boldsymbol{x},\boldsymbol{t})}k(\boldsymbol{t}-\boldsymbol{\theta};\sigma^2)\,d\boldsymbol{t}\right)d\boldsymbol{\omega}\right)d\boldsymbol{y} \qquad (3.30) \\ = \; & \frac{1}{(2\pi)^n}\int_{\mathcal{X}}f(\boldsymbol{y})\left(\int_{\Omega}\int_{\Theta}e^{i\boldsymbol{\omega}^T(\boldsymbol{\tau}(\boldsymbol{x},\boldsymbol{t})-\boldsymbol{y})}k(\boldsymbol{t}-\boldsymbol{\theta};\sigma^2)\,d\boldsymbol{t}\,d\boldsymbol{\omega}\right)d\boldsymbol{y} \\ = \; & \int_{\mathcal{X}}f(\boldsymbol{y})u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y}\,, \qquad (3.31) \end{aligned}$$

where (3.30) uses the ***Parseval*** theorem, and (3.31) uses proposition's assumption (3.29).

□

Proposition 19 *Suppose $f_1=\sum_{k=1}^{p}a_k\phi(\boldsymbol{y};\boldsymbol{x}_k,\delta_k)$, where $\phi(\boldsymbol{x};\boldsymbol{x}_k,\delta_k)=e^{-\frac{\|\boldsymbol{x}-\boldsymbol{x}_k\|^2}{2\delta_k^2}}$. Assume that $u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y})$ is Gaussian in variable $\boldsymbol{y}$. Then the following identity holds.*

$$\begin{aligned} & \int_{\mathcal{X}} f_1(\boldsymbol{y}) u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\ = & \sum_{i=1}^{p} a_i (\frac{\delta_i}{\sqrt{\delta_i^2 + s^2}})^n e^{-\frac{\|\boldsymbol{x}_i - \boldsymbol{\tau}\|^2}{2(\delta_i^2 + s^2)}} . \end{aligned}$$

Proof Sketch

$$\begin{aligned} & \int_{\mathcal{X}} f_1(\boldsymbol{y}) u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \\ = & \int_{\mathcal{X}} f_1(\boldsymbol{y}) k(\boldsymbol{\tau} - \boldsymbol{y}; s^2)\, d\boldsymbol{y} \\ = & \int_{\mathbb{R}^n} \left(\sum_{k=1}^{p} a_k \phi(\boldsymbol{y}; \boldsymbol{x}_k, \delta_k) \right) k(\boldsymbol{\tau} - \boldsymbol{y}; s^2)\, d\boldsymbol{y} \\ = & \sum_{k=1}^{p} a_k \left(\int_{\mathbb{R}^n} \phi(\boldsymbol{y}; \boldsymbol{x}_k, \delta_k) k(\boldsymbol{\tau} - \boldsymbol{y}; s^2)\, d\boldsymbol{y} \right) \\ = & \sum_{k=1}^{p} a_k (\delta_k \sqrt{2\pi})^n \left(\int_{\mathbb{R}^n} k(\boldsymbol{y} - \boldsymbol{x}_k; \delta_k^2) k(\boldsymbol{\tau} - \boldsymbol{y}; s^2)\, d\boldsymbol{y} \right) \\ = & \sum_{k=1}^{p} a_k (\delta_k \sqrt{2\pi})^n \left(\int_{\mathbb{R}^n} \frac{e^{-\frac{\|\boldsymbol{x}_k - \boldsymbol{\tau}\|^2}{2(\delta_k^2 + s^2)}}}{(\sqrt{2\pi(\delta_k^2 + s^2)})^n} k(\boldsymbol{y} - \frac{s^2 \boldsymbol{x}_k + \delta_k^2 \boldsymbol{\tau}}{\delta_k^2 + s^2}; \frac{\delta_k^2 s^2}{\delta_k^2 + s^2}) d\boldsymbol{y} \right) \quad (3.32) \\ = & \sum_{k=1}^{p} a_k (\frac{\delta_k}{\sqrt{\delta_k^2 + s^2}})^n e^{-\frac{\|\boldsymbol{x}_k - \boldsymbol{\tau}\|^2}{2(\delta_k^2 + s^2)}} \left(\int_{\mathbb{R}^n} k(\boldsymbol{y} - \frac{s^2 \boldsymbol{x}_k + \delta_k^2 \boldsymbol{\tau}}{\delta_k^2 + s^2}; \frac{\delta_k^2 s^2}{\delta_k^2 + s^2})\, d\boldsymbol{y} \right) \\ = & \sum_{k=1}^{p} a_k (\frac{\delta_k}{\sqrt{\delta_k^2 + s^2}})^n e^{-\frac{\|\boldsymbol{x}_k - \boldsymbol{\tau}\|^2}{2(\delta_k^2 + s^2)}} , \end{aligned}$$

where in (3.32) we use the Gaussian product result from proposition 0. □

Proposition 20 *The regularized objective function $\tilde{z}(\boldsymbol{\theta}, \boldsymbol{\theta}_0, r, \sigma)$ can be written using transformation kernels as follows.*

$$\begin{aligned} & \tilde{z}(\boldsymbol{\theta}, \boldsymbol{\theta}_0, r, \sigma) \\ = & [\tilde{h}(\,.\,, \boldsymbol{\theta}_0, r) \overset{\Theta}{\circledast} k(\,.\,; \sigma^2)](\boldsymbol{\theta}) \\ = & \int_{\mathcal{X}} \left(k(\boldsymbol{\theta} - \boldsymbol{\theta}_0; r^2 + \sigma^2) f_2(\boldsymbol{x}) \int_{\mathcal{X}} \left(f_1(\boldsymbol{y}) u_{\boldsymbol{\tau}, \frac{r\sigma}{\sqrt{r^2 + \sigma^2}}} (\frac{r^2 \boldsymbol{\theta} + \sigma^2 \boldsymbol{\theta}_0}{r^2 + \sigma^2}, \boldsymbol{x}, \boldsymbol{y}) \right) d\boldsymbol{y} \right) d\boldsymbol{x} . \end{aligned}$$

Proof Sketch For computing $\tilde{z}$, we proceed as below.

$$
\begin{aligned}
& \tilde{z}(\boldsymbol{\theta}, \boldsymbol{\theta}_0, r, \sigma) \\
= \; & [\tilde{h}(\,.\,, \boldsymbol{\theta}_0, r) \star k(\,.\,; \sigma^2)](\boldsymbol{\theta}) \\
= \; & [\Big(\int_{\mathcal{X}} \big(k(\boldsymbol{\theta} - \boldsymbol{\theta}_0; r^2) f_1(\boldsymbol{\tau}(\boldsymbol{x}; \boldsymbol{\theta})) f_2(\boldsymbol{x}) \big) \, d\boldsymbol{x} \Big) \star k(\,.\,; \sigma^2)](\boldsymbol{\theta}) \\
= \; & \int_{\mathcal{X}} \Bigg(f_2(\boldsymbol{x}) [\Big(k(\boldsymbol{\theta} - \boldsymbol{\theta}_0; r^2) f_1(\boldsymbol{\tau}(\boldsymbol{x}; \boldsymbol{\theta})) \Big) \star k(\,.\,; \sigma^2)](\boldsymbol{\theta}) \Bigg) d\boldsymbol{x} \\
= \; & \int_{\mathcal{X}} \Bigg(f_2(\boldsymbol{x}) \int_{\Theta} \Big(k(\boldsymbol{\theta}_0 - \boldsymbol{t}; r^2) f_1(\boldsymbol{\tau}(\boldsymbol{x}; \boldsymbol{t})) k(\boldsymbol{\theta} - \boldsymbol{t}; \sigma^2) \Big) \, d\boldsymbol{t} \Bigg) d\boldsymbol{x} \\
= \; & \int_{\mathcal{X}} \Bigg(f_2(\boldsymbol{x}) \int_{\Theta} \Big(f_1(\boldsymbol{\tau}(\boldsymbol{x}; \boldsymbol{t})) \frac{e^{-\frac{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2}{2(r^2 + \sigma^2)}}}{\Big(\sqrt{2\pi(r^2 + \sigma^2)} \Big)^m} k(\boldsymbol{t} - \frac{\sigma^2 \boldsymbol{\theta}_0 + r^2 \boldsymbol{\theta}}{r^2 + \sigma^2}; \frac{r^2 \sigma^2}{r^2 + \sigma^2}) \Big) \, d\boldsymbol{t} \Bigg) d\boldsymbol{x} \\
= \; & \int_{\mathcal{X}} \Bigg(\frac{e^{-\frac{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2}{2(r^2 + \sigma^2)}}}{\Big(\sqrt{2\pi(r^2 + \sigma^2)} \Big)^m} f_2(\boldsymbol{x}) \int_{\mathcal{X}} \Big(f_1(\boldsymbol{y}) u_{\boldsymbol{\tau}, \frac{r\sigma}{\sqrt{r^2 + \sigma^2}}} (\frac{r^2 \boldsymbol{\theta} + \sigma^2 \boldsymbol{\theta}_0}{r^2 + \sigma^2}, \boldsymbol{x}, \boldsymbol{y}) \Big) \, d\boldsymbol{y} \Bigg) d\boldsymbol{x} \, .
\end{aligned}
$$

Thus, regularized objective function from (3.27) leads to the following result.

$$
\begin{aligned}
& \tilde{z}(\boldsymbol{\theta}, \boldsymbol{\theta}_0, r, \sigma) \\
= \; & [\tilde{h}(\,.\,, \boldsymbol{\theta}_0, r) \overset{\Theta}{\circledast} k(\,.\,; \sigma^2)](\boldsymbol{\theta}) \\
= \; & \int_{\mathcal{X}} \Bigg(k(\boldsymbol{\theta} - \boldsymbol{\theta}_0; r^2 + \sigma^2) f_2(\boldsymbol{x}) \int_{\mathcal{X}} \Big(f_1(\boldsymbol{y}) u_{\boldsymbol{\tau}, \frac{r\sigma}{\sqrt{r^2 + \sigma^2}}} (\frac{r^2 \boldsymbol{\theta} + \sigma^2 \boldsymbol{\theta}_0}{r^2 + \sigma^2}, \boldsymbol{x}, \boldsymbol{y}) \Big) \, d\boldsymbol{y} \Bigg) d\boldsymbol{x} \, .
\end{aligned}
$$

□

3.7.4 Derivation of Affine and Homography Kernels

Proposition 21 *Suppose $n \geq 1$ is some integer and let $t : \mathbb{R}^n \to (\mathbb{R} - \{0\})$. Then for any real $n \times n$ matrix $\boldsymbol{A}^\dagger$ and any real $n \times 1$ vectors $\boldsymbol{b}^\dagger$, $\boldsymbol{x}$, and $\boldsymbol{y}$, the following identity holds:*

$$\begin{aligned}
&\int_{\mathcal{A}}\int_{\mathcal{B}} e^{\frac{i\boldsymbol{\omega}^T\boldsymbol{A}\boldsymbol{x}+i\boldsymbol{\omega}^T\boldsymbol{b}}{t(\boldsymbol{x})}-i\boldsymbol{\omega}^T\boldsymbol{y}} k_\sigma(\boldsymbol{A}-\boldsymbol{A}^\dagger)k_\sigma(\boldsymbol{b}-\boldsymbol{b}^\dagger)\,d\boldsymbol{A}\,d\boldsymbol{b} \\
=\;& k(\frac{\boldsymbol{A}^\dagger\boldsymbol{x}+\boldsymbol{b}^\dagger}{t(\boldsymbol{x})}-\boldsymbol{y};\frac{\sigma^2(1+\|\boldsymbol{x}\|^2)}{t^2(\boldsymbol{x})})\,,
\end{aligned}$$

where $\Omega=\mathcal{B}=\mathbb{R}^n$ *and* $\mathcal{A}=\mathbb{R}^n\times\mathbb{R}^n$.

Proof Sketch We proceed as below,

$$\begin{aligned}
&\int_{\mathcal{A}}\int_{\mathcal{B}} e^{\frac{i\boldsymbol{\omega}^T\boldsymbol{A}\boldsymbol{x}+i\boldsymbol{\omega}^T\boldsymbol{b}}{t(\boldsymbol{x})}-i\boldsymbol{\omega}^T\boldsymbol{y}} k_\sigma(\boldsymbol{A}-\boldsymbol{A}^\dagger)k_\sigma(\boldsymbol{b}-\boldsymbol{b}^\dagger)\,d\boldsymbol{A}\,d\boldsymbol{b} \\
=\;& \int_{\mathcal{A}}\int_{\mathcal{B}} e^{\frac{i\sum_{j=1}^n\sum k=1^n w_j a_{jk}x_k+i\sum_{k=1}^n\omega_k b_k}{t(\boldsymbol{x})}-i\sum_{k=1}^n\omega_k y_k} k_\sigma(\boldsymbol{A}-\boldsymbol{A}^\dagger)k_\sigma(\boldsymbol{b}-\boldsymbol{b}^\dagger)\,d\boldsymbol{A}\,d\boldsymbol{b} \\
=\;& e^{-i\sum_{j=1}^n\omega_j y_j}\Pi_{j=1}^n\Big(\int_{\mathcal{B}_j} e^{\frac{i\omega_j}{t(\boldsymbol{x})}b_j}k_\sigma(b_j-b_j^\dagger)\,db_j\Big) \\
& \Pi_{j=1}^n\Pi_{k=1}^n\Big(\int_{\mathcal{A}_{jk}} e^{\frac{iw_jx_k}{t(\boldsymbol{x})}a_{jk}}k_\sigma(a_{jk}-a_{jk}^\dagger)\,da_{jk}\Big) \\
=\;& e^{-i\sum_{j=1}^n\omega_j y_j}\Pi_{j=1}^n\Big(e^{\frac{i\omega_j}{t(\boldsymbol{x})}b_j^\dagger+\frac{1}{2}\sigma^2(\frac{i\omega_j}{t(\boldsymbol{x})})^2}\Big) \qquad (3.33)\\
& \Pi_{j=1}^n\Pi_{k=1}^n\Big(e^{\frac{iw_jx_k}{t(\boldsymbol{x})}a_{jk}^\dagger+\frac{1}{2}\sigma^2(\frac{iw_jx_k}{t(\boldsymbol{x})})^2}\Big)\,,
\end{aligned}$$

where (3.33) uses the identity $\int_{\mathbb{R}} e^{ax}k_\sigma(x^\dagger-x)\,dx = e^{ax^\dagger+\frac{1}{2}\sigma^2a^2}$. We proceed by factorizing ω_j and ω_j^2 in the exponent as the following.

$$\begin{aligned}
=\;& \int_{\mathcal{A}}\int_{\mathcal{B}} e^{\frac{i\boldsymbol{\omega}^T\boldsymbol{A}\boldsymbol{x}+i\boldsymbol{\omega}^T\boldsymbol{b}}{t(\boldsymbol{x})}-i\boldsymbol{\omega}^T\boldsymbol{y}} k_\sigma(\boldsymbol{A}-\boldsymbol{A}^\dagger)k_\sigma(\boldsymbol{b}-\boldsymbol{b}^\dagger)\,d\boldsymbol{A}\,d\boldsymbol{b} \\
=\;& e^{-i\sum_{j=1}^n\omega_j y_j}\Pi_{j=1}^n\Big(e^{\frac{i\omega_j}{t(\boldsymbol{x})}b_j^\dagger+\frac{1}{2}\sigma^2(\frac{i\omega_j}{t(\boldsymbol{x})})^2}\Big) \\
& \Pi_{j=1}^n\Pi_{k=1}^n\Big(e^{\frac{iw_jx_k}{t(\boldsymbol{x})}a_{jk}^\dagger+\frac{1}{2}\sigma^2(\frac{iw_jx_k}{t(\boldsymbol{x})})^2}\Big) \\
=\;& \Pi_{j=1}^n e^{-i\omega_j y_j+\frac{i\omega_j}{t(\boldsymbol{x})}b_j^\dagger+\frac{1}{2}\sigma^2(\frac{i\omega_j}{t(\boldsymbol{x})})^2+\frac{i(\sum_{k=1}^n a_{jk}^\dagger x_k)}{t(\boldsymbol{x})}w_j+\frac{1}{2}\sigma^2\frac{\sum_{k=1}^n x_k^2}{t^2(\boldsymbol{x})}(iw_j)^2} \\
=\;& \Pi_{j=1}^n e^{i\omega_j\frac{b_j^\dagger-y_j+\sum_{k=1}^n a_{jk}^\dagger x_k}{t(\boldsymbol{x})}-\frac{1}{2}w_j^2\frac{\sigma^2(1+\sum_{k=1}^n x_k^2)}{t^2(\boldsymbol{x})}}
\end{aligned}$$

(3.34)

Now dividing both sides by $(2\pi)^{-n}$ and integrating w.r.t. $\boldsymbol{\omega}$, we obtain the following.

$$
\begin{aligned}
&= (2\pi)^{-n} \int_{\Omega} \int_{\mathcal{A}} \int_{\mathcal{B}} e^{\frac{i\boldsymbol{\omega}^T \boldsymbol{A}\boldsymbol{x} + i\boldsymbol{\omega}^T \boldsymbol{b}}{t(\boldsymbol{x})} - i\boldsymbol{\omega}^T \boldsymbol{y}} k_\sigma(\boldsymbol{A} - \boldsymbol{A}^\dagger) k_\sigma(\boldsymbol{b} - \boldsymbol{b}^\dagger)\, d\boldsymbol{A}\, d\boldsymbol{b}\, d\boldsymbol{\omega} \quad (3.35)\\
&= (2\pi)^{-n} \int_{\Omega} \Pi_{j=1}^n e^{i\omega_j \frac{b_j^\dagger - y_j t(\boldsymbol{x}) + \sum_{k=1}^n a_{jk}^\dagger x_k}{t(\boldsymbol{x})} - \frac{1}{2} w_j^2 \frac{\sigma^2(1 + \sum_{k=1}^n x_k^2)}{t^2(\boldsymbol{x})}}\, d\boldsymbol{\omega}\\
&= \Pi_{j=1}^n \Big(\int_{\Omega_j} (2\pi)^{-1} e^{i\omega_j \frac{b_j^\dagger - y_j t(\boldsymbol{x}) + \sum_{k=1}^n a_{jk}^\dagger x_k}{t(\boldsymbol{x})} - \frac{1}{2} w_j^2 \frac{\sigma^2(1 + \sum_{k=1}^n x_k^2)}{t^2(\boldsymbol{x})}}\, d\omega_j \Big)\\
&= \Pi_{j=1}^n \Big(k(\frac{b_j^\dagger - y_j t(\boldsymbol{x}) + \sum_{k=1}^n a_{jk}^\dagger x_k}{t(\boldsymbol{x})}; \frac{\sigma^2(1 + \sum_{k=1}^n x_k^2)}{t^2(\boldsymbol{x})}) \Big) \quad (3.36)\\
&= k(\frac{\boldsymbol{A}^\dagger \boldsymbol{x} + \boldsymbol{b}^\dagger}{t(\boldsymbol{x})} - \boldsymbol{y}; \frac{\sigma^2(1 + \|\boldsymbol{x}\|^2)}{t^2(\boldsymbol{x})})\,,
\end{aligned}
$$

where (3.36) uses the identity $(2\pi)^{-1} \int_{\mathbb{R}} e^{i\omega x - \frac{\omega^2}{2y}}\, d\omega = k(x; y)$ for $y > 0$.

□

Lemma 22 (Derivation of Affine Kernel) *Suppose $\boldsymbol{x} \in \mathbb{R}^n$, where $n \geq 1$ is some integer. The kernel $u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}^\dagger, \boldsymbol{x}, \boldsymbol{y})$ for the affine transformation $\boldsymbol{\tau}(\boldsymbol{x}) = \boldsymbol{A}^\dagger \boldsymbol{x} + \boldsymbol{b}^\dagger$ is equal to the following expression:*

$$
k\Big(\boldsymbol{A}^\dagger \boldsymbol{x} + \boldsymbol{b}^\dagger - \boldsymbol{y}; \sigma^2(1 + \|\boldsymbol{x}\|^2)\Big)\,,
$$

where $\boldsymbol{A}^\dagger$ is any $n \times n$ real matrix and $\boldsymbol{b}^\dagger$ and $\boldsymbol{y}$ are any $n \times 1$ real vectors.

Proof Sketch By (3.29) from Proposition 14, any u that satisfies the following equation is a kernel for $\boldsymbol{\tau}$.

$$
u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}^\dagger, \boldsymbol{x}, \boldsymbol{y}) \triangleq \frac{1}{(2\pi)^n} \int_{\Omega} \left(\int_{\Theta} e^{i\boldsymbol{\omega}^T (\boldsymbol{\tau}(\boldsymbol{x}, \boldsymbol{\theta}) - \boldsymbol{y})} k_\sigma(\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger)\, d\boldsymbol{\theta} \right)\, d\boldsymbol{\omega}\,.
$$

We proceed with computing u as the following,

$$
\begin{aligned}
& u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}^\dagger, \boldsymbol{x}, \boldsymbol{y})\\
\triangleq\ & \frac{1}{(2\pi)^n} \int_{\Omega} \left(\int_{\Theta} e^{i\boldsymbol{\omega}^T (\boldsymbol{\tau}(\boldsymbol{x}, \boldsymbol{\theta}) - \boldsymbol{y})} k_\sigma(\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger)\, d\boldsymbol{\theta} \right)\, d\boldsymbol{\omega}\\
=\ & \frac{1}{(2\pi)^n} \int_{\Omega} \int_{\mathcal{A}} \int_{\mathcal{B}} e^{i\boldsymbol{\omega}^T (\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} - \boldsymbol{y})} k_\sigma(\boldsymbol{A} - \boldsymbol{A}^\dagger) k_\sigma(\boldsymbol{b} - \boldsymbol{b}^\dagger)\, d\boldsymbol{A}\, d\boldsymbol{b}\, d\boldsymbol{\omega} \quad (3.37)\\
=\ & k\Big(\boldsymbol{A}^\dagger \boldsymbol{x} + \boldsymbol{b}^\dagger - \boldsymbol{y}; \sigma^2(1 + \|\boldsymbol{x}\|^2)\Big)\,,
\end{aligned}
$$

where (3.37) applies Lemma 21 with the particular choice of $t(\boldsymbol{x}) = 1$.

□

Proposition 23 *The following indefinite integral identities hold.*

$$\begin{aligned}
&\forall t \in \mathbb{R}\ ,\ c \in \mathbb{R}\ ,\ p_1 \in \mathbb{R}\quad p_2 \in \mathbb{R}_{++}\ :\\
&\int e^{-p_2 t^2 + p_1 t}\, dt = \frac{1}{2}\sqrt{\frac{\pi}{p_2}} e^{\frac{p_1^2}{4p_2}} \operatorname{erf}(\frac{2p_2 t - p_1}{2\sqrt{p_2}}) + c\\
&\int t e^{-p_2 t^2 + p_1 t}\, dt = \frac{p_1}{4p_2\sqrt{p_2}} e^{\frac{p_1^2}{4p_2}} \sqrt{\pi} \operatorname{erf}(\frac{2p_2 t - p_1}{2\sqrt{p_2}}) - \frac{1}{2p_2} e^{-p_2 t^2 + t p_1} + c\\
&\int t^2 e^{p_1 t - p_2 t^2} dt = \frac{\sqrt{\pi}}{8p_2^2\sqrt{p_2}}(2p_2 + p_1^2) e^{\frac{p_1^2}{4p_2}} \operatorname{erf}(\frac{2p_2 t - p_1}{2\sqrt{p_2}}) - \frac{p_1 + 2p_2 t}{4p_2^2} e^{t p_1 - p_2 t^2} + c.
\end{aligned}$$

Proof Sketch The correctness of these identities can be easily checked by differentiating RHS w.r.t. t and observing that it becomes equal to the integrand of LHS. Remember $\frac{d}{dt}\operatorname{erf}(t) = \frac{2}{\sqrt{\pi}} e^{-t^2}$.

□

Corollary 24 *The following definite integral identities hold.*

$$\begin{aligned}
&\forall \quad t \in \mathbb{R}\ ,\ p_1 \in \mathbb{R}\quad p_2 \in \mathbb{R}_{++}\ :\\
&\int_{\mathbb{R}} e^{-p_2 t^2 + p_1 t}\, dt = \sqrt{\frac{\pi}{p_2}} e^{\frac{p_1^2}{4p_2}}\\
&\int_{\mathbb{R}} t e^{-p_2 t^2 + p_1 t}\, dt = \frac{p_1}{2p_2\sqrt{p_2}} e^{\frac{p_1^2}{4p_2}} \sqrt{\pi}\\
&\int_{\mathbb{R}} t^2 e^{-p_2 t^2 + p_1 t}\, dt = \frac{\sqrt{\pi}}{4p_2^2\sqrt{p_2}}(2p_2 + p_1^2) e^{\frac{p_1^2}{4p_2}}\ .
\end{aligned}$$

Proof Sketch Using the identities for their indefinite counterparts provided in Proposition 23, these definite integrals are easily computed by subtracting their value at the limit $t \to \pm\infty$. Note that $\lim_{t\to\pm\infty} \operatorname{erf}(t) = \pm 1$ and that $\lim_{t\to\pm\infty} f(t)\exp(-p_2 t^2 + p_1 t) = 0$, where $p_2 > 0$ and $f : \mathbb{R} \to \mathbb{R}$ is such that $f(t)$ is a polynomial in t. □

Lemma 25 (Derivation of Homography Kernel) *Suppose $\boldsymbol{x} \in \mathbb{R}^2$. The kernel $u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}^\dagger, \boldsymbol{x}, \boldsymbol{y})$ for the homography transformation $\boldsymbol{\tau}(\boldsymbol{x}) = (\boldsymbol{A}^\dagger \boldsymbol{x} + \boldsymbol{b}^\dagger)(1 + \boldsymbol{c}^{\dagger T}\boldsymbol{x})^{-1}$ is equal to the following expression:*

$$u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}^\dagger, \boldsymbol{x}, \boldsymbol{y}) = qe^{-p}\,,$$

where the auxiliary variables are as below:

$$\begin{aligned}
z_0 &\triangleq \frac{1}{1+\|\boldsymbol{x}\|^2} \\
z_1 &\triangleq 1 + \boldsymbol{x}^T\boldsymbol{c}^\dagger \\
\boldsymbol{v} &\triangleq \boldsymbol{A}^\dagger\boldsymbol{x} + \boldsymbol{b}^\dagger \\
q &\triangleq z_0 \frac{(z_0\|\boldsymbol{x}\|^2\boldsymbol{y}^T\boldsymbol{v} + z_1)^2 + \sigma^2\|\boldsymbol{x}\|^2(1+z_0\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2)}{2\pi\sigma^2(1+z_0\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2)^{\frac{5}{2}}} \\
p &\triangleq \frac{\|z_1\boldsymbol{y} - \boldsymbol{v}\|^2 + z_0\|\boldsymbol{x}\|^2(v_2y_1 - v_1y_2)^2}{2\sigma^2(1+\|\boldsymbol{x}\|^2(1+\|\boldsymbol{y}\|^2))}\,.
\end{aligned}$$

Here $\boldsymbol{A}^\dagger$ is any 2×2 real matrix and $\boldsymbol{b}^\dagger$, $\boldsymbol{c}^\dagger$, and $\boldsymbol{y}$ are any 2×1 real vectors.

Proof Sketch By (3.29) from Proposition 14, any u that satisfies the following equation is a kernel for $\boldsymbol{\tau}$.

$$u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}^\dagger, \boldsymbol{x}, \boldsymbol{y}) \triangleq \frac{1}{(2\pi)^n}\int_\Omega \left(\int_\Theta e^{i\boldsymbol{\omega}^T(\boldsymbol{\tau}(\boldsymbol{x},\boldsymbol{\theta})-\boldsymbol{y})} k_\sigma(\boldsymbol{\theta}-\boldsymbol{\theta}^\dagger)\,d\boldsymbol{\theta}\right) d\boldsymbol{\omega}\,.$$

We proceed with computing u as follows:

$$\begin{aligned}
& u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}^\dagger, \boldsymbol{x}, \boldsymbol{y}) \\
\triangleq\ & \frac{1}{(2\pi)^n}\int_\Omega \left(\int_\Theta e^{i\boldsymbol{\omega}^T(\boldsymbol{\tau}(\boldsymbol{x},\boldsymbol{\theta})-\boldsymbol{y})} k_\sigma(\boldsymbol{\theta}-\boldsymbol{\theta}^\dagger)\,d\boldsymbol{\theta}\right) d\boldsymbol{\omega} \\
=\ & \int_{\mathcal{C}} \Big(\frac{1}{(2\pi)^n}\int_\Omega\int_{\mathcal{A}}\int_{\mathcal{B}} e^{i\boldsymbol{\omega}^T(\frac{\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}}{1+\boldsymbol{c}^T\boldsymbol{x}}-\boldsymbol{y})} k_\sigma(\boldsymbol{A}-\boldsymbol{A}^\dagger) k_\sigma(\boldsymbol{b}-\boldsymbol{b}^\dagger)\,d\boldsymbol{A}\,d\boldsymbol{b}\,d\boldsymbol{\omega}\Big) \\
& \quad k_\sigma(\boldsymbol{c}-\boldsymbol{c}^\dagger)\,d\boldsymbol{c} \\
=\ & \int_{\mathcal{C}} \left(k\Big(\frac{\boldsymbol{A}^\dagger\boldsymbol{x}+\boldsymbol{b}^\dagger}{1+\boldsymbol{c}^T\boldsymbol{x}} - \boldsymbol{y}; \frac{\sigma^2(1+\|\boldsymbol{x}\|^2)}{(1+\boldsymbol{c}^T\boldsymbol{x})^2}\Big) k_\sigma(\boldsymbol{c}-\boldsymbol{c}^\dagger)\right) d\boldsymbol{c} \qquad (3.38) \\
=\ & \int_{\mathcal{C}_2}\int_{\mathcal{C}_1} \Big(k\big(\frac{\boldsymbol{A}^\dagger\boldsymbol{x}+\boldsymbol{b}^\dagger}{1+\boldsymbol{c}^T\boldsymbol{x}} - \boldsymbol{y}; \frac{\sigma^2(1+\|\boldsymbol{x}\|^2)}{(1+\boldsymbol{c}^T\boldsymbol{x})^2}\big) k_\sigma(c_1-c_1^\dagger)dc_1\Big) k_\sigma(c_2-c_2^\dagger)dc_2,
\end{aligned}$$

where (3.38) applies Lemma 21 with the particular choice of $t(\boldsymbol{x}) = 1+\boldsymbol{c}^T\boldsymbol{x}$,

and $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2$ with $\mathcal{C}_1 = \mathcal{C}_2 = \mathbb{R}$.

We continue by first computing the inner integral, aka w.r.t. c_1. To reduce clutter, we introduce the following auxiliary variables which are ***independent*** of c_1.

$$\begin{aligned} \boldsymbol{v} &\triangleq \boldsymbol{A}^\dagger \boldsymbol{x} + \boldsymbol{b}^\dagger \\ s &\triangleq 1 + c_2 x_2 \\ z_0 &\triangleq \frac{1}{\sigma^2(1+\|\boldsymbol{x}\|^2)} \\ z_1 &\triangleq \frac{1}{2\pi\sigma\sqrt{2\pi}}\,. \end{aligned}$$

Now we proceed with integration w.r.t. c_1 as below.

$$\begin{aligned} & \int_{\mathcal{C}_1} k_\sigma(c_1 - c_1^\dagger)\, k(\frac{\boldsymbol{A}^\dagger \boldsymbol{x} + \boldsymbol{b}^\dagger}{1 + \boldsymbol{c}^T\boldsymbol{x}} - \boldsymbol{y}; \frac{\sigma^2(1+\|\boldsymbol{x}\|^2)}{(1+\boldsymbol{c}^T\boldsymbol{x})^2})\, dc_1 \\ = & \int_{\mathcal{C}_1} \frac{1}{\sqrt{2\pi}\sigma} \frac{(1+\boldsymbol{c}^T\boldsymbol{x})^2}{2\pi\sigma^2(1+\|\boldsymbol{x}\|^2)} e^{\frac{-(c_1-c_1^\dagger)^2}{2\sigma^2} - \frac{(1+\boldsymbol{c}^T\boldsymbol{x})^2}{2\sigma^2(1+\|\boldsymbol{x}\|^2)}\|\frac{\boldsymbol{A}^\dagger\boldsymbol{x}+\boldsymbol{b}^\dagger}{1+\boldsymbol{c}^T\boldsymbol{x}} - \boldsymbol{y}\|^2}\, dc_1 \\ = & z_0 z_1 \int_{\mathcal{C}_1} (q_0 + c_1 q_1 + c_1^2 q_2) e^{-p_2 c_1^2 + p_1 c_1 + p_0}\, dc_1 \\ = & z_0 z_1 \sqrt{\frac{\pi}{p_2}} e^{p_0 + \frac{p_1^2}{4p_2}} \Big(q_0 + q_1 \frac{p_1}{2p_2} + q_2 \frac{1}{4p_2^2}(2p_2 + p_1^2) \Big)\,, \end{aligned} \tag{3.39}$$

where (3.39) uses Corollary 24 with the particular choice of p_i and q_i for $i = 0, 1, 2$ as the following. Obviously the following p_2 satisfies $p_2 > 0$. Also not that p_i and q_i are independent of integration variable c_1.

$$\begin{aligned} q_0 &\triangleq s^2 \\ q_1 &\triangleq 2x_1 s \\ q_2 &\triangleq x_1^2 \\ p_0 &\triangleq -\frac{{c_1^\dagger}^2}{2\sigma^2} - \frac{z_0}{2}\|\boldsymbol{v} - s\boldsymbol{y}\|^2 \\ p_1 &\triangleq \frac{c_1^\dagger}{\sigma^2} + z_0\Big(x_1 \boldsymbol{y}^T(\boldsymbol{v} - s\boldsymbol{y})\Big) \\ p_2 &\triangleq \frac{1}{2\sigma^2} + z_0 \frac{x_1^2\|\boldsymbol{y}\|^2}{2}\,. \end{aligned}$$

Combining (3.39) and (3.39) gives the following.

$$\begin{aligned} & u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}^\dagger, \boldsymbol{x}, \boldsymbol{y}) \\ = & \int_{\mathcal{C}_2} z_0 z_1 \sqrt{\frac{\pi}{p_2}} e^{p_0 + \frac{p_1^2}{4p_2}} \Big(q_0 + q_1 \frac{p_1}{2p_2} + q_2 \frac{1}{4p_2^2}(2p_2 + p_1^2) \Big) \, dc_2 \, . \end{aligned}$$

We can compute the above integral in a similar fashion as shown below.

$$\begin{aligned} & u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}^\dagger, \boldsymbol{x}, \boldsymbol{y}) \\ = & z_0 z_1 \sqrt{\frac{\pi}{p_2}} \int_{\mathcal{C}_2} k_\sigma(c_2 - c_2^\dagger) e^{p_0 + \frac{p_1^2}{4p_2}} \Big(q_0 + q_1 \frac{p_1}{2p_2} + q_2 \frac{1}{4p_2^2}(2p_2 + p_1^2) \Big) \, dc_2 \\ = & z_0 z_1 \sqrt{\frac{\pi}{p_2}} \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathcal{C}_2} e^{p_0 + \frac{p_1^2}{4p_2} - \frac{(c_2 - c_2^\dagger)^2}{2}} \Big(q_0 + q_1 \frac{p_1}{2p_2} + q_2 \frac{1}{4p_2^2}(2p_2 + p_1^2) \Big) \, dc_2 \\ = & \frac{z_0 z_1 z_2^2}{|x_2|} \sqrt{\frac{\pi}{p_2}} \int_{\mathcal{C}_2} \frac{1}{\sqrt{2\pi}\sigma} (Q_0 + s Q_1 + s^2 Q_2) e^{-P_2 s^2 + P_1 s + P_0} \, ds \qquad (3.40) \\ = & \frac{z_0 z_1 z_2^2}{\sigma |x_2|} \sqrt{\frac{\pi}{2p_2 P_2}} e^{P_0 + \frac{P_1^2}{4P_2}} \Big(Q_0 + Q_1 \frac{P_1}{2P_2} + Q_2 \frac{1}{4P_2^2}(2P_2 + P_1^2) \Big) , \qquad (3.41) \end{aligned}$$

where (3.40) applies change of variable $s = 1 + x_2 c_2$ to the integral. Note that, $\int_{\mathbb{R}} f(c_2) dc_2 = \text{sign}(x_2) \int_{\mathbb{R}} f((s-1)/x_2) ds/x_2 = 1/|x_2| \int_{\mathbb{R}} f((s-1)/x_2) ds$. Also, (3.41) uses Corollary 24 with the particular choice of z_2, P_i and Q_i for $i = 0, 1, 2$ as the following. Obviously the following P_2 satisfies $P_2 > 0$. Also not that P_i and Q_i are independent of integration variable s.

$$\begin{aligned} z_2 &\triangleq \frac{1}{1 + \sigma^2 x_1^2 z_0 \|\boldsymbol{y}\|^2} \\ Q_0 &\triangleq \frac{1}{z_2} \sigma^2 x_1^2 + x_1^2 (c_1^\dagger + \sigma^2 z_0 x_1 \boldsymbol{y}^T \boldsymbol{v})^2 \\ Q_1 &\triangleq 2 x_1 (c_1^\dagger + \sigma^2 z_0 x_1 \boldsymbol{y}^T \boldsymbol{v}) \\ Q_2 &\triangleq 1 \\ P_0 &\triangleq -\frac{1}{2\sigma^2 x_2^2} (1 + c_2^\dagger x_2)^2 - \frac{z_0 z_2}{2} (\|\boldsymbol{v} - c_1^\dagger x_1 \boldsymbol{y}\|^2 + \sigma^2 z_0 x_1^2 (v_2 y_1 - v_1 y_2)^2) \\ P_1 &\triangleq z_0 z_2 \boldsymbol{y}^T (\boldsymbol{v} - c_1^\dagger x_1 \boldsymbol{y}) + \frac{1}{\sigma^2 x_2^2} (1 + c_2^\dagger x_2) \\ P_2 &\triangleq \frac{1}{2\sigma^2 x_2^2} + \frac{z_0 z_2}{2} \|\boldsymbol{y}\|^2 . \end{aligned}$$

In fact, by plugging in the definitions for z_i, P_i, and Q_i and performing elementary algebraic manipulations, one can write (3.41) more compactly as the following,

$$u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}^\dagger, \boldsymbol{x}, \boldsymbol{y}) = qe^{-p}\,,$$

where the auxiliary variables are as below:

$$\begin{aligned}
z_0 &\triangleq \frac{1}{1+\|\boldsymbol{x}\|^2} \\
z_1 &\triangleq 1 + \boldsymbol{x}^T\boldsymbol{c}^\dagger \\
\boldsymbol{v} &\triangleq \boldsymbol{A}^\dagger\boldsymbol{x} + \boldsymbol{b}^\dagger \\
q &\triangleq z_0 \frac{(z_0\|\boldsymbol{x}\|^2\boldsymbol{y}^T\boldsymbol{v} + z_1)^2 + \sigma^2\|\boldsymbol{x}\|^2(1 + z_0\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2)}{2\pi\sigma^2(1+z_0\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2)^{\frac{5}{2}}} \\
p &\triangleq \frac{\|z_1\boldsymbol{y} - \boldsymbol{v}\|^2 + z_0\|\boldsymbol{x}\|^2(v_2y_1 - v_1y_2)^2}{2\sigma^2(1+\|\boldsymbol{x}\|^2(1+\|\boldsymbol{y}\|^2))}\,.
\end{aligned}$$

□

CHAPTER 4

2D APPLICATIONS

This chapter demonstrates the application of the transformation kernels for 2D alignment tasks. We focus on homography estimation. We explain how the integral transform associated with the homography kernel can be efficiently approximated using the ***Laplace approximation*** scheme. We then show that the proposed method outperforms traditional multi-resolution alignment. We also demonstrate the application of our alignment scheme for 3D reconstruction of an exotic octagonal building.

4.1 Computation with Homography Kernel

As a measure of match between a pair of images, we use the following smoothed correlation function, denoted by $z(\boldsymbol{\theta}, \sigma)$.

$$h(\boldsymbol{\theta}) \triangleq \int_{\mathcal{X}} f_1(\boldsymbol{\tau}(\boldsymbol{x}, \boldsymbol{\theta}))\, f_2(\boldsymbol{x})\, d\boldsymbol{x} \tag{4.1}$$

$$z(\boldsymbol{\theta}, \sigma) \triangleq [h \star k(\,\cdot\,, \sigma^2)](\boldsymbol{\theta})\,, \tag{4.2}$$

where $\boldsymbol{\theta} \triangleq [\, vec(\boldsymbol{A}_{n\times n})\,,\, \boldsymbol{b}_{n\times 1}\,,\, \boldsymbol{c}_{n\times 1}\,]$ and $\boldsymbol{\tau}(\boldsymbol{x}, \boldsymbol{\theta}) \triangleq \frac{1}{1+\boldsymbol{c}^T\boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$. In Chapter 3, it was shown that z can be equivalently computed using the following blur operator.

$$z(\boldsymbol{\theta}, \sigma) \tag{4.3}$$

$$\triangleq [h \star k(\,\cdot\,, \sigma^2)](\boldsymbol{\theta}) \tag{4.4}$$

$$= \int_{\mathcal{X}} \big(f_2(\boldsymbol{x}) [f_1(\boldsymbol{\tau}(\boldsymbol{x}, .)) \star k(\,\cdot\,, \sigma^2)](\boldsymbol{\theta}) \big)\, d\boldsymbol{x} \tag{4.5}$$

$$= \int_{\mathcal{X}} \left(f_2(\boldsymbol{x}) \left(\int_{\mathcal{X}} f_1(\boldsymbol{y}) u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} \right) \right) d\boldsymbol{x}\,, \tag{4.6}$$

where $u_{\boldsymbol{\tau},\sigma}$ is the associated blur kernel. In our 2D experiments, we specifically focus on homography alignment. Chapter 3 derived the following expression for the ***homography kernel*** .

$$
\begin{aligned}
u_{\boldsymbol{\tau},\sigma} &\triangleq q(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y},\sigma)\; e^{-p(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y},\sigma)} \\
p &\triangleq \frac{\|\gamma_1\boldsymbol{y}-\boldsymbol{v}\|^2+\gamma_0\|\boldsymbol{x}\|^2(v_2y_1-v_1y_2)^2}{2\sigma^2(1+\|\boldsymbol{x}\|^2(1+\|\boldsymbol{y}\|^2))} \\
q &\triangleq \gamma_0\frac{(\gamma_0\|\boldsymbol{x}\|^2\boldsymbol{y}^T\boldsymbol{v}+\gamma_1)^2+\sigma^2\|\boldsymbol{x}\|^2(1+\gamma_0\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2)}{2\pi\sigma^2(1+\gamma_0\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2)^{\frac{5}{2}}} \\
\gamma_0 &\triangleq \frac{1}{1+\|\boldsymbol{x}\|^2} \\
\gamma_1 &\triangleq 1+\boldsymbol{c}^T\boldsymbol{x} \\
\boldsymbol{v} &\triangleq \boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}
\end{aligned}
$$

We use piecewise constant forms to represent images (i.e. each pixel is modeled by a continuous intensity whose value remains constant across that pixel). Thus, the integral transform (4.6) can be computed as follows,

$$
\int_{\mathcal{X}} f_1(\boldsymbol{y})u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y} \tag{4.7}
$$

$$
= \sum_{i=1}^{W}\sum_{j=1}^{H} F_1(i,j)\int_{\underline{y_i}}^{\overline{y_i}}\int_{\underline{y_j}}^{\overline{y_j}} u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y}\,, \tag{4.8}
$$

where F_1 is defined on $\{1,2,\ldots,M\}\times\{1,2,\ldots,N\}$; M and N are number of pixels (width and height of the image). Here we define $\underline{y_i} \triangleq \frac{i-1}{M-1}$, $\overline{y_i} \triangleq \frac{i}{M-1}$, $\underline{y_j} \triangleq \frac{j-1}{N-1}$, $\overline{y_j} \triangleq \frac{j}{N-1}$. Using such piecewise constant image model, the only integration required for computing the integral transform is $\int_{\mathcal{X}_{ij}} u_{\boldsymbol{\tau},\sigma}(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y}$.

We use the ***Laplace approximation scheme*** for computing an approximate value of the integral transform (4.6). Consider the function $q(\boldsymbol{y})\,e^{-p(\boldsymbol{y})}$. When p and q grow in similar order in $\boldsymbol{y}$, e.g. both are polynomial or rational functions in $\boldsymbol{y}$, then the behavior of $q(\boldsymbol{y})\,e^{-p(\boldsymbol{y})}$ is dominated at the local maxima of $-p(\boldsymbol{y})$ (see [52] for details). The idea of the Laplace method is to approximate $q(\boldsymbol{y})e^{-p(\boldsymbol{y})}$ around each of these local maxima by a Gaussian function, which is easy to integrate. In particular, when $-p(\boldsymbol{y})$ has only one

bounded maximizer $\boldsymbol{y}^*$, then the approximate becomes as below,

$$u \triangleq q(\boldsymbol{y})\, e^{-p(\boldsymbol{y})} \approx q(\boldsymbol{y}^*) e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{y}^*)^T \nabla^2 p(\boldsymbol{y}^*)(\boldsymbol{y}-\boldsymbol{y}^*)}\,. \tag{4.9}$$

In the following proposition, we show that when $1+\boldsymbol{c}^T\boldsymbol{x} \neq 0$, then $-p$ has a unique maximizer w.r.t variable $\boldsymbol{y}$ attained at $\boldsymbol{y}^* = \frac{\boldsymbol{A}\boldsymbol{x}+\boldsymbol{x}}{1+\boldsymbol{c}^T\boldsymbol{x}}$.

Proposition 26 *Suppose $1+\boldsymbol{c}^T\boldsymbol{x} \neq 0$. The $-p$ in the homography kernel has a unique maximizer w.r.t variable $\boldsymbol{y}$, which is attained at the following point $\boldsymbol{y}^*$,*

$$\boldsymbol{y}^* = \frac{\boldsymbol{A}\boldsymbol{x}+\boldsymbol{x}}{1+\boldsymbol{c}^T\boldsymbol{x}}\,. \tag{4.10}$$

Proof Sketch The exponent $-p$ is as below,

$$-p = \frac{\|\gamma_1\boldsymbol{y}-\boldsymbol{v}\|^2 + \gamma_0\|\boldsymbol{x}\|^2(v_2y_1 - v_1y_2)^2}{2\sigma^2(1+\|\boldsymbol{x}\|^2(1+\|\boldsymbol{y}\|^2))} \tag{4.11}$$

The stationary points of $-p$ can be found by zero crossing its gradient with respect to $\boldsymbol{y}$. Doing so leads to the following pair of points.

$$\boldsymbol{y}_1^* = \frac{\boldsymbol{v}}{\gamma_1} \tag{4.12}$$

$$\boldsymbol{y}_2^* = -\boldsymbol{v}\frac{\gamma_1}{\gamma_0\|\boldsymbol{v}\|^2\|\boldsymbol{x}\|^2} \tag{4.13}$$

We now show that only $\boldsymbol{y}_1^*$ is a maximizer of $-p$. This can be done by examining the eigenvalues of the Hessian of $-p$, evaluated at the stationary points. Denote the eigenvalues of the Hessian by λ and Λ. Then, at each of the $\boldsymbol{y}_1^*$ and $\boldsymbol{y}_2^*$, the eigenvalues are as the following,

$$\lambda_1 = -\frac{\gamma_1^4}{\sigma^2(\|\boldsymbol{v}\|^2\|\boldsymbol{x}\|^2 + \gamma_1^2(1+\|\boldsymbol{x}\|^2))} \tag{4.14}$$

$$\Lambda_1 = -\frac{\gamma_1^2(\gamma_1^2 + \gamma_0\|\boldsymbol{v}\|^2\|\boldsymbol{x}\|^2)}{\sigma^2(\|\boldsymbol{v}\|^2\|\boldsymbol{x}\|^2 + \gamma_1^2(1+\|\boldsymbol{x}\|^2))} \tag{4.15}$$

$$\lambda_2 = 0 \tag{4.16}$$

$$\Lambda_2 = \frac{\|\boldsymbol{v}\|^4\|\boldsymbol{x}\|^4}{\sigma^2(1+\|\boldsymbol{x}\|^2)^2(\|\boldsymbol{v}\|^2\|\boldsymbol{x}\|^2 + \gamma_1^2(1+\|\boldsymbol{x}\|^2))}\,. \tag{4.17}$$

Since both eigenvalues at $\boldsymbol{y}_1^*$ are negative, the exponent $-p$ is strictly

concave at that point and thus attains its maximizer there. However, both eigenvalues at $\boldsymbol{y}_2^*$ are non-negative, and thus $-p$ is convex at that point and thus attains its minimizer there.

□

On the other hand, let's refer to $\nabla^2 p(\boldsymbol{y}^*)$ as $\boldsymbol{H}$, which is a 2×2 matrix. It is easy to check that the elements of $\boldsymbol{H}$ have the following form,

$$h_{11} = \frac{\gamma_1^2(\gamma_1^2 + v_2^2\gamma_0\|\boldsymbol{x}\|^2)}{\sigma^2(\frac{\gamma_1^2}{\gamma_0} + \|\boldsymbol{x}\|^2\|\boldsymbol{v}\|^2)} \tag{4.18}$$

$$h_{22} = \frac{\gamma_1^2(\gamma_1^2 + v_1^2\gamma_0\|\boldsymbol{x}\|^2)}{\sigma^2(\frac{\gamma_1^2}{\gamma_0} + \|\boldsymbol{x}\|^2\|\boldsymbol{v}\|^2)} \tag{4.19}$$

$$h_{12} = h_{21} = -\frac{\gamma_0 v_1 v_2}{\sigma^2(1 + \frac{1}{\|\boldsymbol{x}\|^2} + \frac{\|\boldsymbol{v}\|^2}{\gamma_1^2})}\,, \tag{4.20}$$

Plugging (4.18) into (4.9) leads to the following approximation for the kernel u.

$$\begin{aligned}
u &\triangleq q(\boldsymbol{y})e^{-p(\boldsymbol{y})} && (4.21)\\
&\approx q(\boldsymbol{y}^*)e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{y}^*)^T\boldsymbol{H}(\boldsymbol{y}-\boldsymbol{y}^*)} && (4.22)\\
&= q(\boldsymbol{y}^*)\frac{2\pi}{\sqrt{\det(\boldsymbol{H})}}\frac{\sqrt{\det(\boldsymbol{H})}}{2\pi}e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{y}^*)^T\boldsymbol{H}(\boldsymbol{y}-\boldsymbol{y}^*)} && (4.23)\\
&= q(\boldsymbol{y}^*)\frac{2\pi}{\sqrt{\det(\boldsymbol{H})}}K(\boldsymbol{y}-\boldsymbol{y}^*;\boldsymbol{H}^{-1}) && (4.24)\\
&= \left(1 + \frac{\sigma^2\|\boldsymbol{x}\|^2}{\gamma_1^2 + \gamma_0\,\|\boldsymbol{x}\|^2\,\|\boldsymbol{v}\|^2}\right)K(\boldsymbol{y}-\boldsymbol{y}^*;\boldsymbol{H}^{-1})\,, && (4.25)
\end{aligned}$$

where K denotes a Gaussian kernel. Note that this approximation preserves the limit behavior, i.e. it approaches Dirac's δ function as $\sigma \to 0$.

Now (4.25) is a Gaussian form with covariance matrix $\boldsymbol{C} \triangleq \boldsymbol{H}^{-1}$.

$$\boldsymbol{C} = \frac{\sigma^2\|\boldsymbol{x}\|^2}{\gamma_1^4}\begin{bmatrix} v_1^2 + \frac{\gamma_1^2}{\gamma_0\|\boldsymbol{x}\|^2} & v_1v_2 \\ v_1v_2 & v_2^2 + \frac{\gamma_1^2}{\gamma_0\|\boldsymbol{x}\|^2} \end{bmatrix}. \tag{4.26}$$

However, it is still difficult to integrate (4.25) unless the covariance matrix is diagonal. Thus, we further approximate (4.26) by a diagonal matrix. We

do this in the simplest possible way, i.e. setting off-diagonals to zeros[1],

$$\hat{\boldsymbol{C}} = \frac{\sigma^2\|\boldsymbol{x}\|^2}{\gamma_1^4}\begin{bmatrix} v_1^2 + \frac{\gamma_1^2}{\gamma_0\|\boldsymbol{x}\|^2} & 0 \\ 0 & v_2^2 + \frac{\gamma_1^2}{\gamma_0\|\boldsymbol{x}\|^2} \end{bmatrix}. \tag{4.27}$$

This approximation allows to decouple integrals with respect to y_1 and y_2 and thus to obtain the following closed form:

$$\begin{aligned} & \int_{\underline{y_1}}^{\overline{y_1}} \int_{\underline{y_2}}^{\overline{y_2}} K(\boldsymbol{y} - \boldsymbol{y}^*; \hat{\boldsymbol{C}})\, d\boldsymbol{y} && (4.28) \\ = \; & \Big(\int_{\underline{y_1}}^{\overline{y_1}} k(y_1 - y_1^*; \hat{c}_{11})\, dy_1 \Big) \Big(\int_{\underline{y_2}}^{\overline{y_2}} k(y_2 - y_2^*; \hat{c}_{22})\, dy_2 \Big) && (4.29) \\ = \; & \frac{1}{4}\Big(\operatorname{erf}(\frac{\overline{y_1} - y_1^*}{\sqrt{2\hat{c}_{11}}}) - \operatorname{erf}(\frac{\underline{y_1} - y_1^*}{\sqrt{2\hat{c}_{11}}}) \Big) \Big(\operatorname{erf}(\frac{\overline{y_2} - y_2^*}{\sqrt{2\hat{c}_{22}}}) - \operatorname{erf}(\frac{\underline{y_2} - y_2^*}{\sqrt{2\hat{c}_{22}}}) \Big) && (4.30) \end{aligned}$$

Thus, the integral of the kernel u can approximated as follows,

$$\begin{aligned} & \int_{\underline{y_1}}^{\overline{y_1}} \int_{\underline{y_2}}^{\overline{y_2}} u\, d\boldsymbol{y} && (4.31) \\ \approx \; & \Big(1 + \frac{\sigma^2\|\boldsymbol{x}\|^2}{\gamma_1^2 + \gamma_0\,\|\boldsymbol{x}\|^2\;\|\boldsymbol{v}\|^2}\Big) && (4.32) \\ & \frac{1}{4}\Big(\operatorname{erf}(\frac{\overline{y_1} - y_1^*}{\sqrt{2\hat{c}_{11}}}) - \operatorname{erf}(\frac{\underline{y_1} - y_1^*}{\sqrt{2\hat{c}_{11}}}) \Big) \Big(\operatorname{erf}(\frac{\overline{y_2} - y_2^*}{\sqrt{2\hat{c}_{22}}}) - \operatorname{erf}(\frac{\underline{y_2} - y_2^*}{\sqrt{2\hat{c}_{22}}}) \Big) && (4.33) \end{aligned}$$

Thus (4.33) provides a closed form approximation for integration in (4.8). Although (4.33) is indeed closed form, it involves the erf function, which is usually very slow. We therefore approximate erf function by the following form [53],

$$\begin{aligned} \operatorname{erf}(x) \; & \approx \; \operatorname{sign}(x)\sqrt{1 - \exp(-x^2\,\frac{\frac{4}{\pi} + \alpha x^2}{1 + \alpha x^2})} && (4.34) \\ \alpha \; & \triangleq \; 8\frac{\pi - 3}{3\pi(4 - \pi)} && (4.35) \end{aligned}$$

[1]Admittedly, this is not the best way for approximating the covariance with a diagonal matrix. Better alternatives could be investigated in the future.

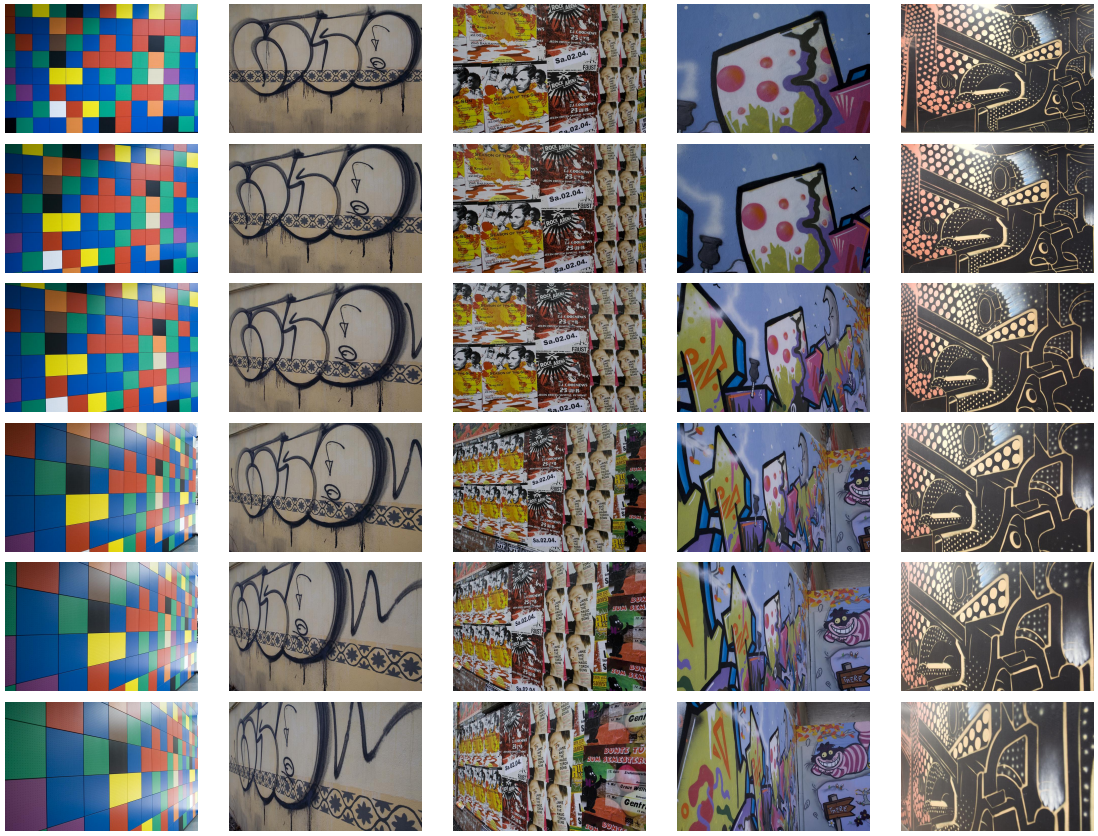


Figure 4.1: Dataset provided by [1] consisting of five planar scenes, each having six different views of increasingly dramatic perspective effect.

4.2 Quantitative Results on Planar Scenes

Dataset. We evaluate the performance of the proposed alignment scheme against traditional Gaussian blurring and no blurring at all. We use the images provided by [1] (see figure 4.1). This dataset consists of five ***planar scenes*** , each having six different views of increasingly dramatic perspective effect. The planar nature of the scenes makes it very suitable for homography alignment purpose.

Alignment. For the proposed method, we use the homography kernel. The goal is to maximize the correlation between a pair of views by transforming one to the other. The local maximization in our alignment algorithm (see Chapter 3), as well as that of other methods used here for comparison, is achieved by a block coordinate ascent method with a naive line search. The

block coordinate ascent is performed by partitioning the 8 parameters of homography to three classes, those that comprise $\boldsymbol{A}$, $\boldsymbol{b}$ and $\boldsymbol{c}$. This improves numerical stability because the sensitivity of parameters within each partition are similar.

Parameters Setup. Pixel coordinates were normalized to range in $[-1, 1]$. Images f_1 and f_2 were converted to grayscale and were subtracted by their joint mean (i.e. $(\bar{f}_1 + \bar{f}_2)/2$, where $\bar{f}_1$ is the average intensity of f_1) as a preprocessing step. The sequence of σ (for both the proposed kernel and Gaussian kernel) starts from $\sigma = 0.1$, and is multiplied by $2/3$ in each iteration of algorithm until it falls below 0.0001. The initial transformation $\boldsymbol{\theta}_0$ was set to the identity, i.e. $\boldsymbol{A}_0 = \boldsymbol{I}$ and $\boldsymbol{b}_0 = \boldsymbol{c}_0 = \boldsymbol{0}$. Since the initial σ is not large and the images lack significant areas of symmetry, no regularization was used.

Performance. The performance of these methods is summarized in figure 4.2-bottom. Each plot corresponds to one of the scenes in the dataset. For each scene, there is one rectified view that is used as f_1. The rest of five views, indexed from 1 to 5, in increasing order of complexity[2] are used as f_2. The vertical axis in the plots indicates the normalized correlation coefficient (NCC) between f_2 and transformed $\tilde{f}_1$. It can clearly be observed that while Gaussian blur sometimes does a little bit better than no blur, the proposed smoothing scheme leads to a much higher NCC value[3].

4.3 Qualitative Results on 3D Reconstruction

Here, we present the application of the proposed alignment method within a bigger task of 3D reconstruction. The previous experiments used scenes with planar structure so that the perspective distorted images could be aligned by a ***single homography*** . However, in a realistic scenario, barely the entire scene consists of a single planar surface. Here, we work with a real application of 3D reconstruction of an exotic octagonal building.

[2]Here the complexity of the view is referred to how drastic the homography transformation is, in order to bring it to the rectified view.

[3]The code for reproducing our results is available at `http://perception.csl.illinois.edu/smoothing`.

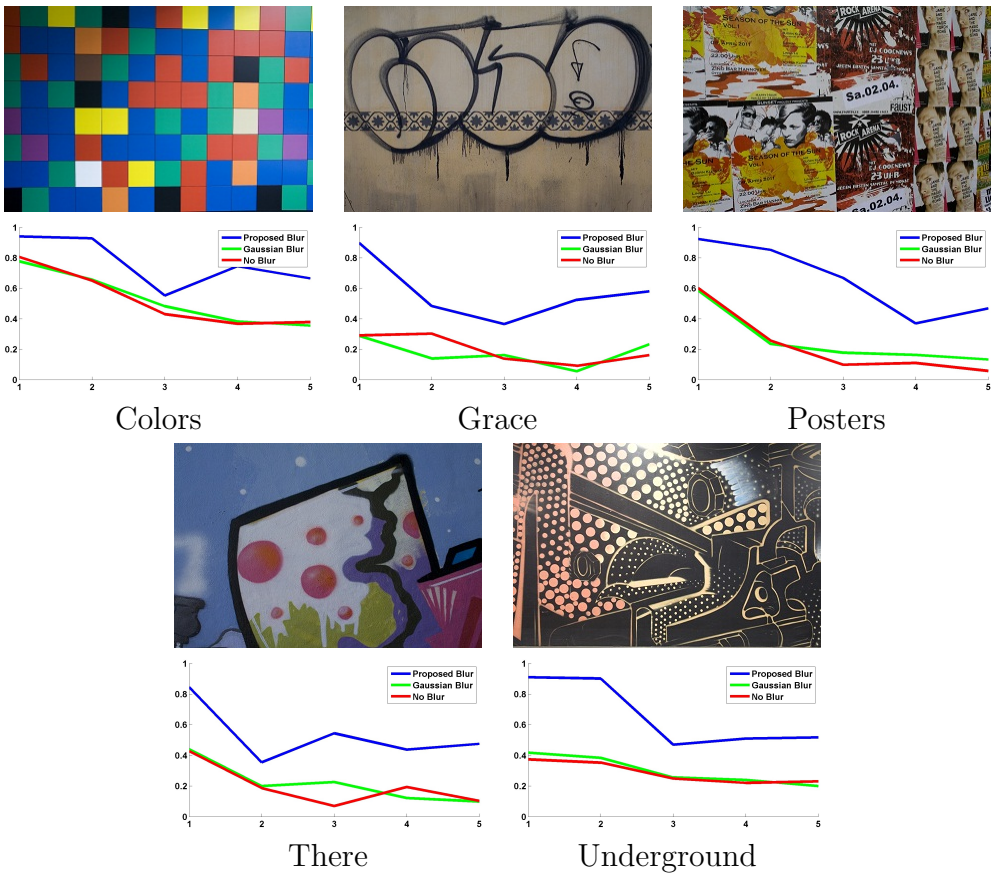


Figure 4.2: **Top**: Representative rectified views from the dataset provided in [1]. **Bottom**: NCC value after alignment. Horizontal axis is the view index (increasing in complexity) of the scene. Four views are used for each scene, each one being as f_2 and compared against f_1, which is a rectified view in the dataset.

The 3D Reconstruction Pipeline. Detailed description of our 3D reconstruction system is beyond the scope of this dissertation and can be found in our paper [31]. However, we provide a brief overview of the pipeline here. We use only eight ***uncalibrated and widely separated*** images for the full reconstruction of the building. Each of the images covers a pair of adjacent facades as shown in Figure 4.3. We arrange the sequence of images so that matching of common facades is only performed between consecutive images. We first segment each image into piecewise planar regions. Once segmented, each region can be rectified using a single homography due to its planar structure. Both segmentation and rectification are guided by the low-rank texture assumption of the facades. Using such rectified textural regions, solv-

Figure 4.3: A pair of matched regions from the same facade with different partial occlusion.

ing wide-baseline correspondence between pair of successive images becomes better conditioned (say by a similarity match). Thus, now we can obtain dense pixel-wise match across pairs of facades using the proposed alignment technique. Finally, for global consistency, we use a scheme similar to the bundle adjustment in conventional structure from motion (SFM). Figure 4.4 shows the reconstructed full 3D model as well as the recovered camera poses. As one can see, despite unknown calibration, partial occlusion, large baselines, our method is able to recover a very precise and complete 3D model of the building.

Dense Correspondence by Alignment. The above segmentation procedure provides a good estimate for the relative location of the facades and their rectified texture (see Figure 4.5 (a) and (b)). However, each segmented region may not share the same ***location and scale*** in different images. Therefore, we need to refine their location and scale in order to obtain ***precise point-wise matching*** between images. Similar to previous section, we use cross correlation to measure the similarity between the two regions. We ***smooth this objective*** along translation and scaling dimensions of the transformation space. We then solve the derived optimization problem using the continuation procedure.

Comparison with Feature Matching. An example of final matching results between two images are given in Figure 4.5. As a comparison, in Figure 4.5 (e), we illustrate the difficulty of applying the classical SIFT matching technique [2] to the urban scenes with repetitive or symmetric patterns. Point-wise matching of low-rank regions outperforms SIFT in this scenario

because the texture segmentation enables us to perform accurate region-based matching rather than using local points or edges. See Figure 4.3 for additional example of the matched facades.

Comparison with other SFM Systems. It is difficult to make a fair comparison between the proposed approach and other structure from motion (SFM) methods, since the large baselines and rich symmetry makes other methods fail. In fact, we tested our sequences on almost all publicly available SFM packages such as Bundler [54], SFM-SIFT [4] (which combines Torr's SFM toolbox [55] with SIFT feature detector [2]), FIT3D [56], and Voodoo Camera Tracker [5]. All these packages report errors related to their inability of establishing meaningful correspondence across the views.

[4] http://homepages.inf.ed.ac.uk/s0346435/projects/sfm/sfm_sift.html
[5] http://www.digilab.uni-hannover.de/docs/manual.html

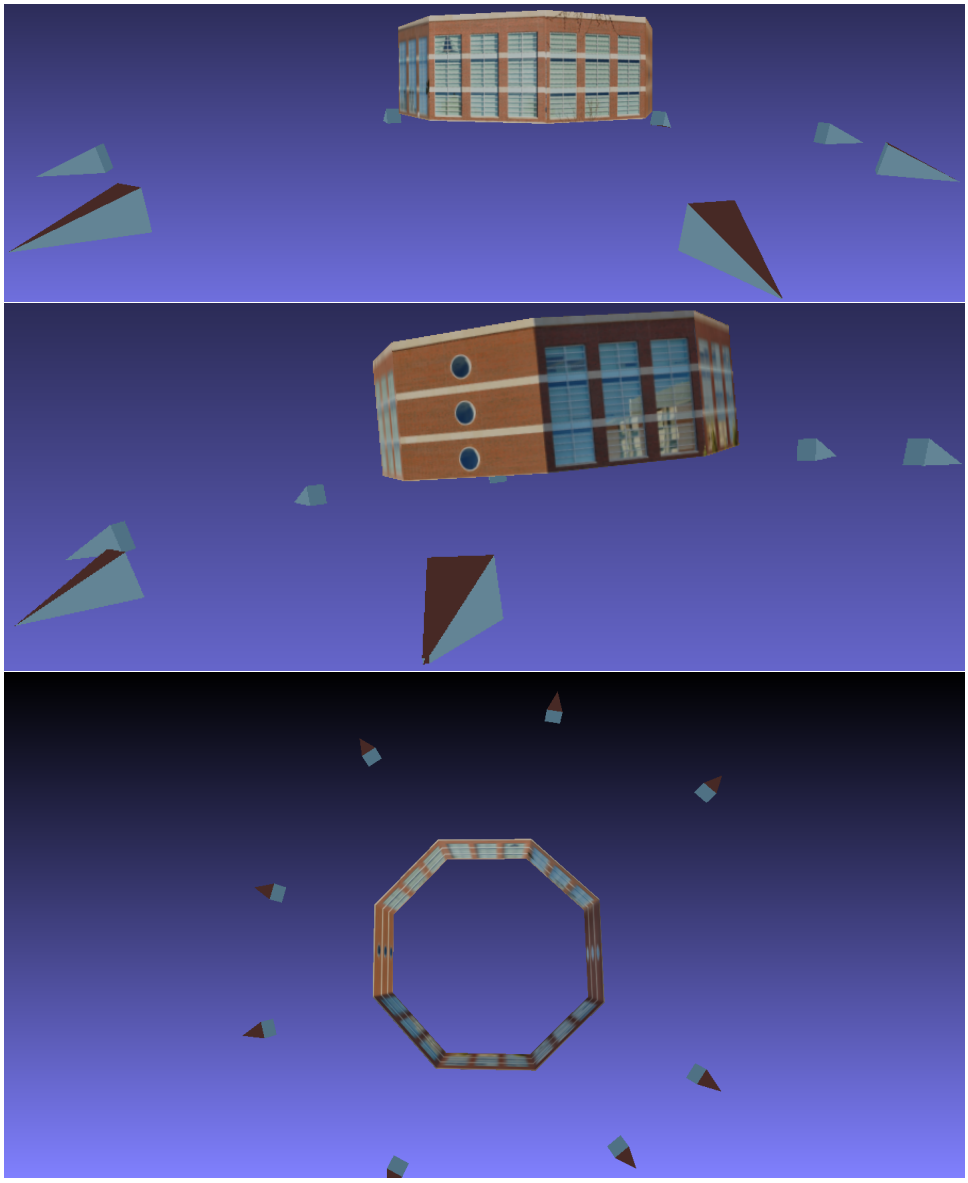


Figure 4.4: Frontal and top views of the recovered building. Each pyramid shows the estimated location of a camera.

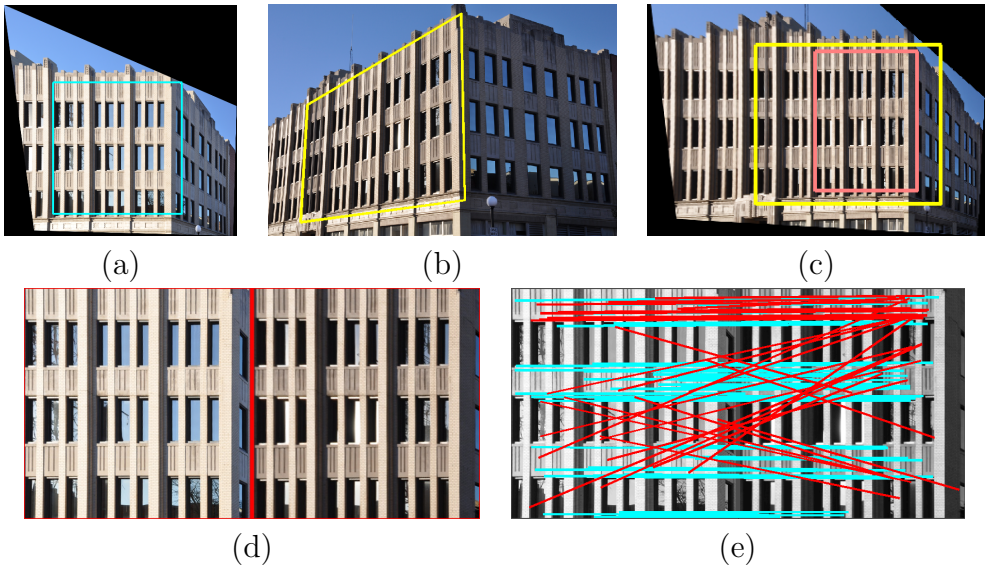


Figure 4.5: (a) Segmented and rectified facade. (b),(c) Same task from a different view. (c) Segmentation result refined to the orange box by matching. (d) Point-wise match between two regions of the facades using our method. (e) Feature-point matching result of the two rectified regions by SIFT [2], with red lines indicating mismatches.

CHAPTER 5

3D APPLICATIONS

This chapter demonstrates the application of the transformation kernels for 3D alignment tasks. Here the goal is to perform ***affine*** alignment of 3D point cloud data. The standard algorithm for point cloud alignment is ***iterative closest point (ICP)*** algorithm [57]. The idea is to simply alternate between creating a correspondence between pair of points (of the two clouds) and local optimization of geometric transformation between the corresponding points. This algorithm is known to work only when good initialization is provided. However, it is susceptible to get stuck in local minima if the transformation is drastic and no prior knowledge is available for initialization. We show that, using the proposed smoothing and continuation scheme, we can outperform ICP. The construction of the smoothed objective in this chapter will use the affine kernel[1] developed in Chapter 2.

5.1 Problem Formulation

Given two sets of points $\mathcal{P} = \{\boldsymbol{p}_i\}_{i=1}^m$ and $\mathcal{Q} = \{\boldsymbol{q}_i\}_{i=1}^n$, where each point belongs to $\mathbb{R}^d$, and d is the dimension of the ambient space, e.g. $d = 3$ for 3D point clouds. We assume that the mean points in each set is zero [2], i.e. $\sum_{i=1}^m \boldsymbol{p}_i = \sum_{j=1}^n \boldsymbol{q}_j = \mathbf{0}$. We refer to $\mathcal{P}$ as the model, and $\mathcal{Q}$ as the data. The data can possibly cover the model partially, e.g. it may have holes. Consider a family of geometric transformations parameterized by $\boldsymbol{\theta} \in \mathbb{R}^t$ and denote it by the map $\boldsymbol{\tau} : \mathbb{R}^d \times \mathbb{R}^t \to \mathbb{R}^d$. We define the optimal transformation as the one which minimizes the following cost function,

[1] Nice thing about affine transformation is that, unlike homography that we derived its kernel only for 2D, the former's kernel is derived for any arbitrary dimension.

[2] If that is not the case, we can always subtract $\frac{1}{m}\sum_{i=1}^m \boldsymbol{p}_i$ from points in $\mathcal{P}$ and $\frac{1}{n}\sum_{j=1}^n \boldsymbol{q}_j$ from $\mathcal{Q}$ to make the mean point in these points equal to zero.

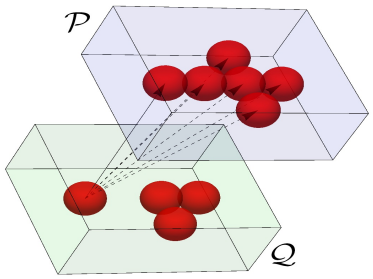


Figure 5.1: Each point in $\mathcal{Q}$ is put into correspondence with one of the points in $\mathcal{P}$.

$$(\boldsymbol{\theta}^*, \boldsymbol{c}^*) = \arg\min_{\boldsymbol{\theta},\boldsymbol{c}} \sum_{i=1}^{m}\sum_{j=1}^{n} c_{i,j}\|\boldsymbol{\tau}(\boldsymbol{p}_i,\boldsymbol{\theta}) - \boldsymbol{q}_j\|^2 \tag{5.1}$$

$$\text{s.t.} \quad \forall\, j \in \{1,\ldots,n\} \quad \sum_{i=1}^{m} c_{i,j} = 1 \tag{5.2}$$

$$\forall\, i \in \{1,\ldots,m\}\, \forall\, j \in \{1,\ldots,n\} \quad c_{i,j} \in \{0,1\}\,. \tag{5.3}$$

The auxiliary variables $\{c_{i,j}\}$ determine the correspondence among the point pairs in $\mathcal{P}$ and $\mathcal{Q}$. Specifically, this formulation requires each point in the data $\mathcal{Q}$, indexed by j, to have a corresponding point[3], indexed by i, in the model $\mathcal{P}$ (see figure 5.1). By concatenating all $c_{i,j}$ into a long vector, we obtain a vector of length $m\,n$, which we denote by $\boldsymbol{c}$.

We approximate this optimization task by an unconstrained one using ***quadratic penalty*** method. The idea is that each equality constraint of form $f(\boldsymbol{\theta},\boldsymbol{c}) = 0$ can be treated as a penalty by adding $f^2(\boldsymbol{\theta},\boldsymbol{c})$ to the objective function points. Note that the discrete constraint $c_{i,j} \in \{0,1\}$ can be equivalently expressed by a continuous equality constraint of form $c_{i,j}(1-c_{i,j}) = 0$. Thus, the approximate objective function becomes as the following,

[3]Since we assume $\mathcal{P}$ is the complete model, regardless of $\mathcal{Q}$ being complete or partial, each point in $\mathcal{Q}$, there should be a corresponding point in $\mathcal{P}$.

$$(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{c}}) = \arg\min_{\boldsymbol{\theta},\boldsymbol{c}} h(\boldsymbol{\theta}, \boldsymbol{c}) \tag{5.4}$$

$$h(\boldsymbol{\theta}, \boldsymbol{c}) \triangleq \epsilon\Big(\sum_{i=1}^{m}\sum_{j=1}^{n} c_{i,j}\|\boldsymbol{\tau}(\boldsymbol{p}_i, \boldsymbol{\theta}) - \boldsymbol{q}_j\|^2\Big) \tag{5.5}$$

$$+\sum_{j=1}^{n}(1 - \sum_{i=1}^{m} c_{i,j})^2 + \sum_{i=1}^{m}\sum_{j=1}^{n} c_{i,j}^2(1 - c_{i,j})^2\,, \tag{5.6}$$

where $\epsilon > 0$ is a small number.

5.2 Smoothing

Instead of directly minimizing h in (5.4), we propose to minimize a smoothed version of h. We construct the smoothed h by Gaussian convolution in the following way, and refer to it as z. We assume that $\boldsymbol{\tau}$ is an ***affine transformation*** , i.e. $\boldsymbol{\tau}\big(\boldsymbol{p}; (\boldsymbol{A}, \boldsymbol{b})\big) \triangleq \boldsymbol{A}\boldsymbol{p} + \boldsymbol{b}$. Since adding constants (w.r.t. optimization variables) does not affect the minimizer, by doing that and some abuse of notation, we can express the smoothed z as below (see section 5.7 for derivation with help of ***affine transformation kernel***).

$$z(\boldsymbol{\theta}, \boldsymbol{c}; \sigma) \triangleq \Big[\Big([h(\,.\,,\,.\,) \star k(\,.\,; \sigma^2)]\,(\boldsymbol{c})\Big) \star k(\,.\,; \sigma^2)\Big]\,(\boldsymbol{\theta}) \tag{5.7}$$

$$= \epsilon\sum_{i=1}^{m}\sum_{j=1}^{n} c_{i,j}\big(\|\boldsymbol{\tau}(\boldsymbol{p}_i, \boldsymbol{\theta}) - \boldsymbol{q}_j\|^2 + 3\sigma^2(1 + \|\boldsymbol{p}_i\|^2)\big) \tag{5.8}$$

$$+\sum_{j=1}^{n}(1 - \sum_{i=1}^{m} c_{i,j})^2 \tag{5.9}$$

$$+\sum_{i=1}^{m}\sum_{j=1}^{n}(c_{i,j} - 1)^2 c_{i,j}^2 + 6\sigma^2(c_{i,j} - \frac{1}{2})^2\,. \tag{5.10}$$

Here the inner k is a multivariate Gaussian with $m\,n$ variables and covariance of $\sigma^2\boldsymbol{I}$. The outer k is also a multivariate Gaussian, but with t variables and again covariance of $\sigma^2\boldsymbol{I}$.

5.3 Gradients

Gradients of z w.r.t. optimization variables $(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c})$ are needed by the minmization algorithm. They are provided in the following.

$$\frac{\partial z}{\partial \boldsymbol{A}} = 2\epsilon \sum_{i=1}^{m}\sum_{j=1}^{n} c_{i,j}\Big(\big(\boldsymbol{A}\boldsymbol{p}_i + \boldsymbol{b} - \boldsymbol{q}_j \big)\boldsymbol{p}_i^T \Big) \tag{5.11}$$

$$\frac{\partial z}{\partial \boldsymbol{b}} = 2\epsilon \sum_{i=1}^{m}\sum_{j=1}^{n} c_{i,j}\Big(\boldsymbol{A}\boldsymbol{p}_i + \boldsymbol{b} - \boldsymbol{q}_j \Big) \tag{5.12}$$

$$\frac{\partial z}{\partial c_{i,j}} = \epsilon\Big(\|\boldsymbol{A}\boldsymbol{p}_i + \boldsymbol{b} - \boldsymbol{q}_j\|^2 + 3\sigma^2(1 + \|\boldsymbol{p}_i\|^2) \Big) - 2 + 2\sum_{k=1}^{m} c_{k,j} \tag{5.13}$$

$$+2\Big(c_{i,j}(c_{i,j} - 1)(2c_{i,j} - 1) + 6\sigma^2(c_{i,j} - \frac{1}{2}) \Big)\,. \tag{5.14}$$

5.4 Asymptotic Properties

Consider the following definitions.

$$\boldsymbol{P} \triangleq \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{p}_i\boldsymbol{p}_i^T \tag{5.15}$$

$$\boldsymbol{U} \triangleq \frac{1}{m\,n}\sum_{i=1}^{m}\sum_{j=1}^{n} \boldsymbol{q}_i\boldsymbol{p}_i^T\,. \tag{5.16}$$

It can be shown that the asymptotic minimizer of z, i.e. when $\sigma \to \infty$, has the following form when $\epsilon \to 0$. See the Section 5.7 for the derivation.

$$\boldsymbol{A}^* = \boldsymbol{U}\,\boldsymbol{P}^{-1} \tag{5.17}$$

$$\boldsymbol{b}^* = \boldsymbol{0} \qquad \boldsymbol{c}^* = \frac{1}{2}\,. \tag{5.18}$$

5.5 Illustrative Example

We present an illustrative example to show how smoothing and path-following may help escaping local minima for the point cloud alignment problem. Due

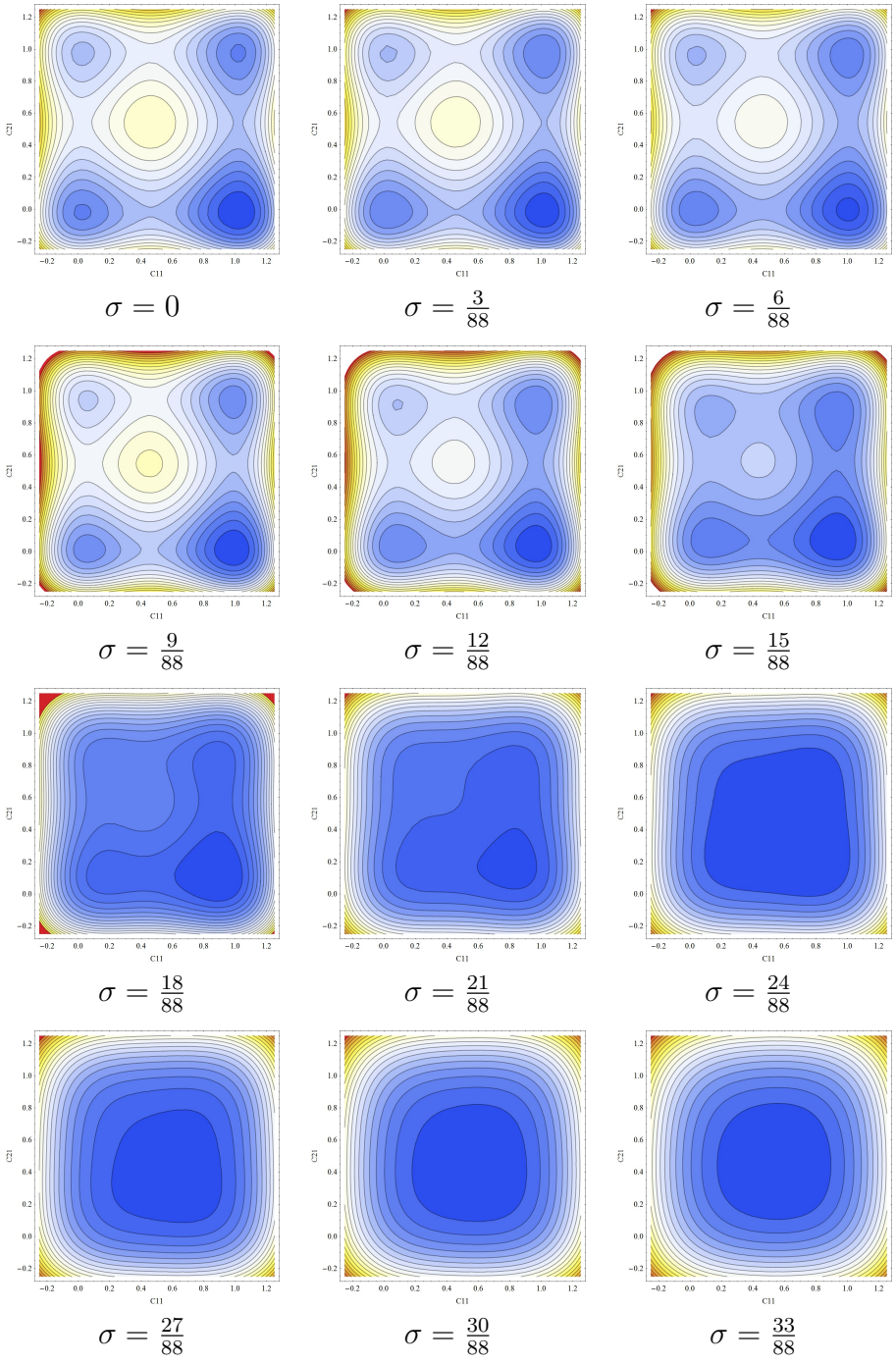


Figure 5.2: Optimization landscape for minimizing the function (5.19). The spectrum from blue to red indicates small to large values.

to 2D visualization constraints for the optimization landscape, we restrict the task to only having two optimization variables. This is explained in the following.

Suppose $d = 1$, i.e. the points are on the real axis $\mathbb{R}$. Consider the sets $\mathcal{P} = \{p_1, p_2\}$ and $\mathcal{Q} = \{q_1, q_2\}$, where $p_1 = q_1 = -1$ and $p_2 = q_2 = 1$. Since the points are already aligned, we know the optimal transformation is just the identity map. Thus we only look at the alignment objective when the transformation is set to the identity and seek for the optimal correspondence. In addition, we know that at the optimum, $c^*_{1,2} = 1 - c_{1,1}$ and $c^*_{2,2} = 1 - c_{2,1}$. Plugging that into the objective function as well, we obtain the following objective in only two variables $c_{1,1}$ and $c_{2,1}$. Obviously, since the points are already aligned, the correspondence correspondence must associate p_1 to q_1, and p_2 to q_2, which implies $c^*_{1,1} = c^*_{2,2} = 1$ and $c^*_{1,2} = c^*_{2,1} = 0$. However, we try to find the optimal $c_{1,1}$ and $c_{2,1}$ via optimization and see if the smoothing can help finding the optimal solution.

$$
\begin{aligned}
& h(\boldsymbol{c}) \\
&= \epsilon\Big(\sum_{i=1}^{m}\sum_{j=1}^{n} c_{i,j}(p_i - q_j)^2\Big) \\
&\quad + \sum_{j=1}^{n}(1 - \sum_{i=1}^{m} c_{i,j})^2 + \sum_{i=1}^{m}\sum_{j=1}^{n} c_{i,j}^2(1 - c_{i,j})^2 \\
&= \epsilon\Big(c_{1,1}(p_1 - q_1)^2 + c_{2,1}(p_2 - q_1)^2 + (1 - c_{1,1})(p_1 - q_2)^2 + (1 - c_{2,1})(p_2 - q_2)^2\Big) \\
&\quad + (1 - c_{1,1})^2 c_{1,1}^2 + (1 - c_{2,1})^2 c_{2,1}^2 \\
&= 4\epsilon(1 - c_{1,1} + c_{2,1}) + (1 - c_{1,1})^2 c_{1,1}^2 + (1 - c_{2,1})^2 c_{2,1}^2 \,.
\end{aligned}
$$

Let's choose $\epsilon = 0.01$. The Gaussian convolution of the objective h w.r.t. variables $c_{1,1}$ and $c_{2,1}$ leads to the following function.

$$
\begin{aligned}
& z(c_{1,1}, c_{2,1}; \sigma) \qquad (5.19) \\
= \; & \frac{1}{25}\Big(1 + c_{1,1}(25(-1 + c_{1,1})^2 c_{1,1} - 1) + c_{2,1}(1 + 25(c_{2,1} - 1)^2 c_{2,1}) \\
& \quad + 50\sigma^2 + 150((c_{1,1} - 1)c_{1,1} + (c_{2,1} - 1)c_{2,1})\sigma^2 + 150\sigma^4\Big) \,.
\end{aligned}
$$

The optimization landscape is shown in figure (5.2). Observe that, at the

Algorithm 3 Point Cloud Affine Alignment by Gaussian Smoothing

Input: Zero mean model cloud $\mathcal{P}$ and data cloud $\mathcal{Q}$, sequence $\sigma_1 > \sigma_2 > \cdots > \sigma_N > 0$, small $\epsilon > 0$, $c_{i,j} = \frac{1}{2}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$.
$\boldsymbol{A}_0 = \boldsymbol{U}\,\boldsymbol{P}^{-1}$
$\boldsymbol{b}_0 = \boldsymbol{0}$
$\boldsymbol{c}_0 = \{\frac{1}{2}\}^{m\times n}$
$k = 1$
repeat
  $(\boldsymbol{A}_k, \boldsymbol{b}_k, \boldsymbol{c}_k) = \arg\min_{(\boldsymbol{A},\boldsymbol{b},\boldsymbol{c})} \epsilon \sum_{i=1}^{m}\sum_{j=1}^{n} c_{i,j}\big(\|\boldsymbol{\tau}(\boldsymbol{p}_i, \boldsymbol{\theta}) - \boldsymbol{q}_j\|^2 + 3\sigma_k^2(1 + \|\boldsymbol{p}_i\|^2)\big) + \sum_{j=1}^{n}(1 - \sum_{i=1}^{m} c_{i,j})^2 + \sum_{i=1}^{m}\sum_{j=1}^{n}(c_{i,j} - 1)^2 c_{i,j}^2 + 6\sigma_k^2(c_{i,j} - \frac{1}{2})^2$ // Local minimization initialized at $(\boldsymbol{A}_{k-1}, \boldsymbol{b}_{k-1}, \boldsymbol{c}_{k-1})$.
  $k = k + 1$
until $k > N$
Output: $(\boldsymbol{A}_k, \boldsymbol{b}_k, \boldsymbol{c}_k)$

non-smoothed function (i.e. when $\sigma = 0$), besides the global minimum near $(c_{1,1}, c_{2,1}) = (1, 0)$, there are three local minima near $(0, 1)$, $(0, 0)$ and $(1, 1)$. However by starting from a large enough σ, we obtain a convex landscape whose minimizer is near $(c_{1,1}, c_{2,1}) = (\frac{1}{2}, \frac{1}{2})$, as anticipated by the asymptotic minimizer result. Following the path of minimizer, originated form the asymptotic minimizer, as σ is shrunk down to 0, the method leads to the global minimizer of the actual function near $(c_{1,1}, c_{2,1}) = (\frac{1}{2}, \frac{1}{2})$.

5.6 Algorithm & Results

The Algorithm 3 shows the procedure for affine alignment by Gaussian smoothing and path following. Note that the minimization inside the loop is done locally using an initial point.

We apply this algorithm to some of the 3D objects in Stanford's dataset [58, 59]. For each object, we rotate the model by n degrees along all three x, y and z axes, where n varies between 30 degrees to 90 degrees, in steps of 15 degrees. The alignment results shown in Figures 5.3, 5.4, and 5.5 are quite encouraging. Specially for bunny and Buddha objects, the smoothing method works much better than ICP. For Dragon object, ICP and the proposed method both do bad, although the proposed method seems slightly better.

5.7 Proofs

5.7.1 Derivation of Gaussian Smoothed Objective

We compute the inner convolution as follows.

$$
\begin{aligned}
& [h(\boldsymbol{\theta}\,,\,.\,) \star k(\,.\,;\sigma^2)]\,(\boldsymbol{c}) && (5.20)\\
= \;& \epsilon\Big(\sum_{i=1}^{m}\sum_{j=1}^{n} c_{i,j}\|\boldsymbol{\tau}(\boldsymbol{p}_i,\boldsymbol{\theta}) - \boldsymbol{q}_j\|^2\Big) && (5.21)\\
& + m\,n\,\sigma^2 + \sum_{j=1}^{n}(1-\sum_{i=1}^{m} c_{i,j})^2 && (5.22)\\
& + 3\,m\,n\,\sigma^4 + \sum_{i=1}^{m}\sum_{j=1}^{n}(c_{i,j}-1)^2 c_{i,j}^2 + \sigma^2(1+6c_{i,j}(c_{i,j}-1)) && (5.23)
\end{aligned}
$$

Since adding constant terms, i.e. those that do not depend on $c_{i,j}$ and $\boldsymbol{\theta}$, does not affect the minimizer of the optimization, with some abuse of notation, we do that and express the result of (5.20) as follows.

$$
\begin{aligned}
& [h(\boldsymbol{\theta}\,,\,.\,) \star k(\,.\,;\sigma^2)]\,(\boldsymbol{c}) && (5.24)\\
= \;& \epsilon\Big(\sum_{i=1}^{m}\sum_{j=1}^{n} c_{i,j}\|\boldsymbol{\tau}(\boldsymbol{p}_i,\boldsymbol{\theta}) - \boldsymbol{q}_j\|^2\Big) && (5.25)\\
& + \sum_{j=1}^{n}(1-\sum_{i=1}^{m} c_{i,j})^2 && (5.26)\\
& + \sum_{i=1}^{m}\sum_{j=1}^{n}(c_{i,j}-1)^2 c_{i,j}^2 + 6\sigma^2(c_{i,j}-\frac{1}{2})^2 && (5.27)
\end{aligned}
$$

Observe that Gaussian smoothing of h w.r.t. $\boldsymbol{c}$ leads to the same h plus a regularization term $6\sigma^2(c_{i,j}-\frac{1}{2})^2$, which enhances convexity of the objective with larger choices of σ. We now apply the convolution w.r.t. $\boldsymbol{\theta}$. We assume that $\boldsymbol{\tau}$ is an ***affine transformation*** , i.e. $\boldsymbol{\tau}\big(\boldsymbol{p};(\boldsymbol{A},\boldsymbol{b})\big) \triangleq \boldsymbol{A}\boldsymbol{p}+\boldsymbol{b}$.

$$
\begin{aligned}
z(\boldsymbol{\theta}, \boldsymbol{c}; \sigma) \;&\triangleq\; \left[\Big([h(\,.\,,\,.\,) \star k(\,.\,; \sigma^2)]\,(\boldsymbol{c}) \Big) \star k(\,.\,; \sigma^2) \right](\boldsymbol{\theta}) \\
&= \Bigg[\Big(\epsilon \big(\sum_{i=1}^{m} \sum_{j=1}^{n} c_{i,j} \|\boldsymbol{\tau}(\boldsymbol{p}_i, \boldsymbol{\theta}) - \boldsymbol{q}_j\|^2 \big) \\
&\qquad + \sum_{j=1}^{n} (1 - \sum_{i=1}^{m} c_{i,j})^2 \qquad (5.28) \\
&\qquad + \sum_{i=1}^{m} \sum_{j=1}^{n} (c_{i,j} - 1)^2 c_{i,j}^2 + 6\sigma^2 (c_{i,j} - \frac{1}{2})^2 \Big) \star k(\,.\,; \sigma^2) \Bigg](\boldsymbol{\theta}) \\
&= \epsilon \big(\sum_{i=1}^{m} \sum_{j=1}^{n} c_{i,j} \int_{\mathbb{R}^3} \|\boldsymbol{r} - \boldsymbol{q}_j\|^2 k(\boldsymbol{\tau}(\boldsymbol{p}_i, \boldsymbol{\theta}) - \boldsymbol{r}; \sigma^2 (1 + \|\boldsymbol{p}_i\|^2)) d\boldsymbol{r} \qquad (5.29) \\
&\qquad + \sum_{j=1}^{n} (1 - \sum_{i=1}^{m} c_{i,j})^2 \\
&\qquad + \sum_{i=1}^{m} \sum_{j=1}^{n} (c_{i,j} - 1)^2 c_{i,j}^2 + 6\sigma^2 (c_{i,j} - \frac{1}{2})^2 \\
&= \epsilon \sum_{i=1}^{m} \sum_{j=1}^{n} c_{i,j} \big(\|\boldsymbol{\tau}(\boldsymbol{p}_i, \boldsymbol{\theta}) - \boldsymbol{q}_j\|^2 + 3\sigma^2 (1 + \|\boldsymbol{p}_i\|^2) \big) \qquad (5.30) \\
&\qquad + \sum_{j=1}^{n} (1 - \sum_{i=1}^{m} c_{i,j})^2 \\
&\qquad + \sum_{i=1}^{m} \sum_{j=1}^{n} (c_{i,j} - 1)^2 c_{i,j}^2 + 6\sigma^2 (c_{i,j} - \frac{1}{2})^2 \,,
\end{aligned}
$$

where (5.29) uses the ***transformation kernel*** for the affine map introduced in Chapter 3. This kernel allows writing the high dimensional convolution w.r.t. $\boldsymbol{\theta}$ equivalently by a d-dimensional integral transform.

5.7.2 Derivation of the Asymptotic Minimizer

This objective look asymptotically (when $\sigma \to \infty$) looks as the following,

$$
\lim_{\sigma \to \infty} z(\boldsymbol{\theta}, \boldsymbol{c}; \sigma) \;=\; 3\sigma^2 \sum_{i=1}^{m} \sum_{j=1}^{n} \Big(\epsilon c_{i,j} (1 + \|\boldsymbol{p}_i\|^2) + 2 (c_{i,j} - \frac{1}{2})^2 \Big) \,. \qquad (5.31)
$$

It can be seen that the choice of $\boldsymbol{\theta}$ has no effect on the minimizing $\boldsymbol{c}$. The objective (5.31) is ***convex*** in variable $\boldsymbol{c}$, thus it has a unique minimizer. The minimizer is obtained by zero crossing the gradient of (5.31) w.r.t. $\boldsymbol{c}$, as shown below.

$$c^*_{i,j} = \frac{1}{2} - \frac{1}{4}\epsilon(1 + \|\boldsymbol{p}_i\|^2)\,. \tag{5.32}$$

By plugging in this value into the actual objective (5.30), we obtain a minimization task which only depends on $\boldsymbol{\theta}$. By collecting all the remaining terms that are constant terms w.r.t. $\boldsymbol{\theta}$ into a term called t, this objective looks as the following,

$$\lim_{\sigma\to\infty} z(\boldsymbol{\theta}, \boldsymbol{c}^*; \sigma) \;=\; \epsilon \sum_{i=1}^{m}\sum_{j=1}^{n} c^*_{i,j}\|\boldsymbol{\tau}(\boldsymbol{p}_i, \boldsymbol{\theta}) - \boldsymbol{q}_j\|^2 \;+\; t\,. \tag{5.33}$$

This function is ***convex*** in $\boldsymbol{\theta}$ when $\boldsymbol{\tau}$ is an affine transformation. Therefore, it has a unique minimizer w.r.t. $\boldsymbol{\theta}$. If we multiply this function by any positive constant, and in particularly by $(m\,n\,\epsilon)^{-1}$, the minimizer $\boldsymbol{\theta}^*$, does not change. Thus, we proceed as below. First we provide some definitions.

$$\boldsymbol{P} \;\triangleq\; \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{p}_i\boldsymbol{p}_i^T \tag{5.34}$$

$$\boldsymbol{U} \;\triangleq\; \frac{1}{m\,n}\sum_{i=1}^{m}\sum_{j=1}^{n} \boldsymbol{q}_i\boldsymbol{p}_i^T\,. \tag{5.35}$$

We can now continue as the following,

$$\lim_{\sigma\to\infty} (m\,n\,\epsilon)^{-1}\, z(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}^*; \sigma) \tag{5.36}$$

$$= \frac{1}{m\,n}\sum_{i=1}^{m}\sum_{j=1}^{n} c^*_{i,j}\|\boldsymbol{A}\boldsymbol{p}_i + \boldsymbol{b} - \boldsymbol{q}_j\|^2 \;+\; \frac{t}{m\,n\,\epsilon} \tag{5.37}$$

$$= \frac{1}{m\,n}\sum_{i=1}^{m}\sum_{j=1}^{n} c^*_{i,j}(\boldsymbol{A}\boldsymbol{p}_i + \boldsymbol{b} - \boldsymbol{q}_j)^T(\boldsymbol{A}\boldsymbol{p}_i + \boldsymbol{b} - \boldsymbol{q}_j) \;+\; \frac{t}{m\,n\,\epsilon}\,. \tag{5.38}$$

Due to its convexity, the unique minimizer of this function can be obtained

by zero crossing its derivative[4] w.r.t. $\boldsymbol{A}$ and $\boldsymbol{b}$ as shown below. We also use the fact that when $\epsilon \to 0$, then $c_{i,j} \to \frac{1}{2}$.

$$\Rightarrow \quad \frac{\partial}{\partial \boldsymbol{A}} \lim_{\epsilon \to 0} \lim_{\sigma \to \infty} (m\, n\, \epsilon)^{-1}\, z(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}^*; \sigma) = \boldsymbol{A}\boldsymbol{P} - \boldsymbol{U} \tag{5.42}$$

$$\frac{\partial}{\partial \boldsymbol{b}} \lim_{\epsilon \to 0} \lim_{\sigma \to \infty} (m\, n\, \epsilon)^{-1}\, z(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}^*; \sigma) = \boldsymbol{b}\,. \tag{5.43}$$

By zero crossing these equations and solving them in $\boldsymbol{A}$ and $\boldsymbol{b}$, it follows that,

$$\boldsymbol{A}^* = \boldsymbol{U}\,\boldsymbol{P}^{-1} \tag{5.44}$$

$$\boldsymbol{b}^* = \boldsymbol{0}\,. \tag{5.45}$$

[4]Suppose $\boldsymbol{u}$ and $\boldsymbol{v}$ are $n \times 1$ matrices and $\boldsymbol{A}$ is a $n \times n$ matrix. Then we have the following identities for derivative of a scalar w.r.t. the matrix $\boldsymbol{A}$.

$$\frac{\partial \boldsymbol{u}^T \boldsymbol{A} \boldsymbol{v}}{\partial \boldsymbol{A}} = \boldsymbol{u}\boldsymbol{v}^T \tag{5.39}$$

$$\frac{\partial \boldsymbol{u}^T \boldsymbol{A}^T \boldsymbol{v}}{\partial \boldsymbol{A}} = \boldsymbol{v}\boldsymbol{u}^T \tag{5.40}$$

$$\frac{\partial \boldsymbol{u}^T \boldsymbol{A}^T \boldsymbol{A} \boldsymbol{u}}{\partial \boldsymbol{A}} = 2\boldsymbol{A}\boldsymbol{u}\boldsymbol{u}^T\,. \tag{5.41}$$

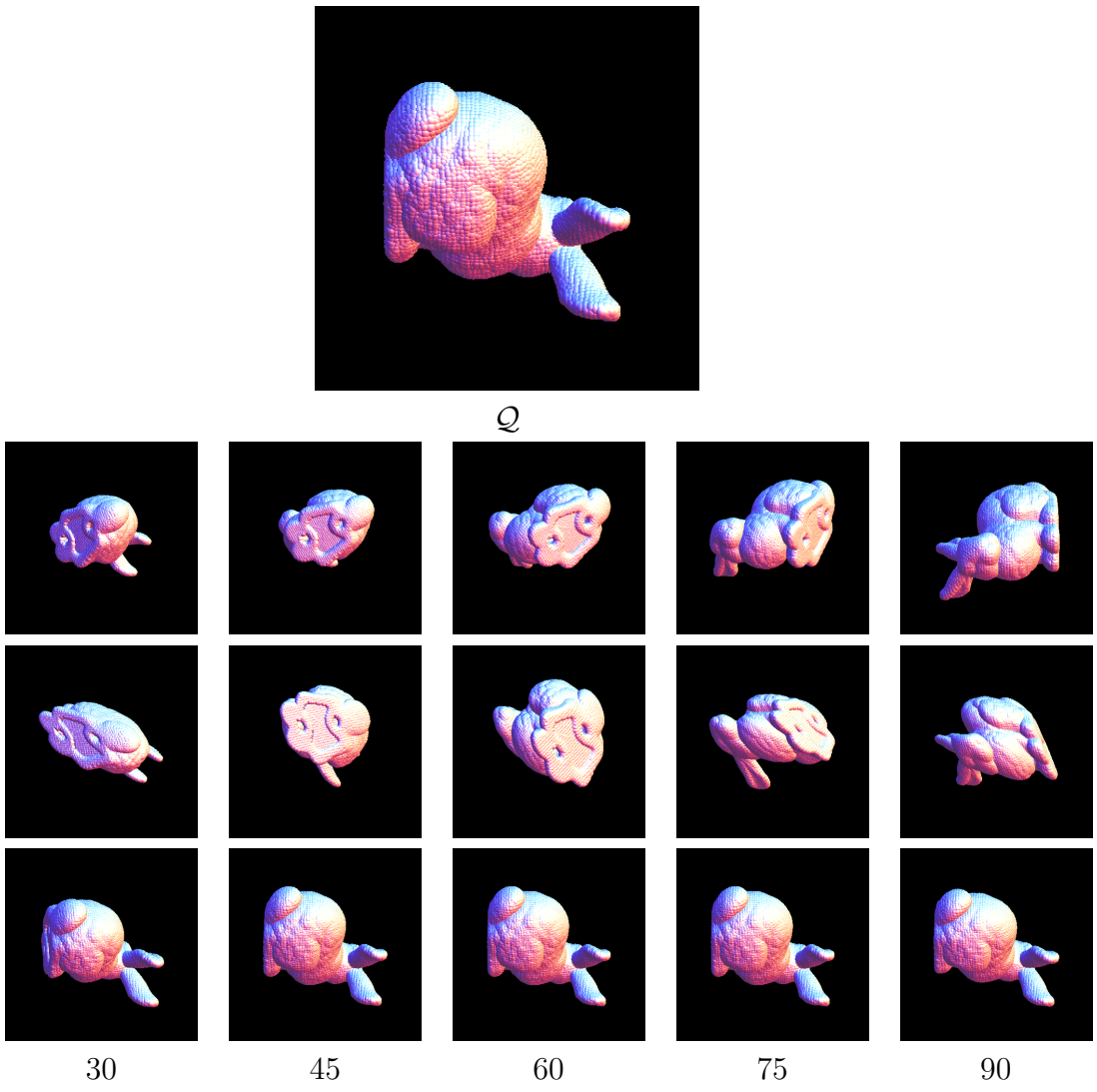


Figure 5.3: Top Row : Input $\mathcal{P}$, which is a rotated version of $\mathcal{Q}$. Middle Row : Transformed $\mathcal{P}$ to match $\mathcal{Q}$ using ICP. Bottom Row: Transformed $\mathcal{P}$ to match $\mathcal{Q}$ using proposed method.

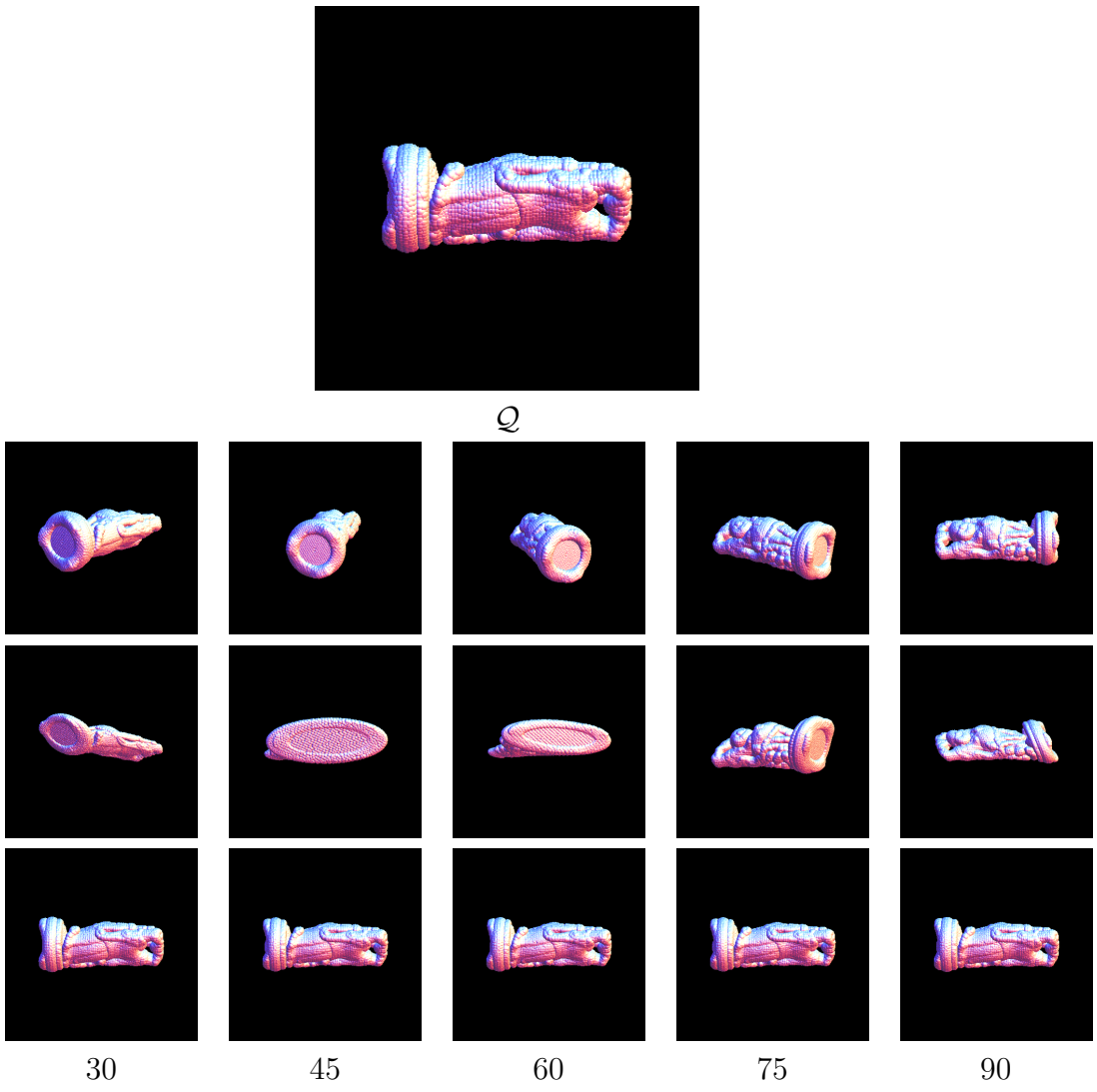


Figure 5.4: Top Row : Input $\mathcal{P}$, which is a rotated version of $\mathcal{Q}$. Middle Row : Transformed $\mathcal{P}$ to match $\mathcal{Q}$ using ICP. Bottom Row: Transformed $\mathcal{P}$ to match $\mathcal{Q}$ using proposed method.

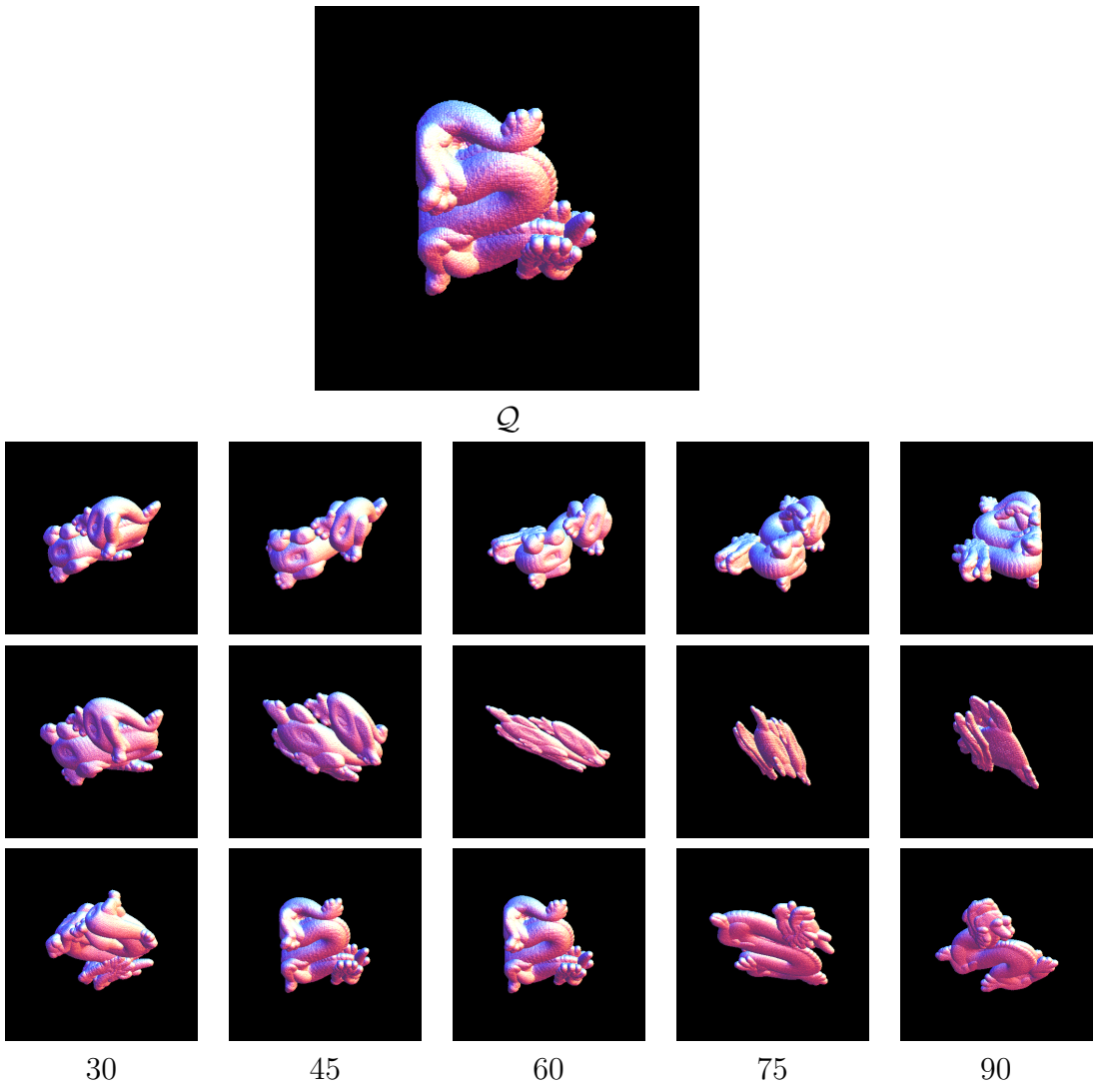


Figure 5.5: Top Row : Input $\mathcal{P}$, which is a rotated version of $\mathcal{Q}$. Middle Row : Transformed $\mathcal{P}$ to match $\mathcal{Q}$ using ICP. Bottom Row: Transformed $\mathcal{P}$ to match $\mathcal{Q}$ using proposed method.

CHAPTER 6

CONCLUSION & FUTURE DIRECTIONS

6.1 Continuation

This dissertation rigorously investigated some of the fundamental properties of the smoothing technique. We presented a formal definition for asymptotic convexity. We gave closed form and derivative free expressions for checking asymptotic convexity as well as deriving the asymptotic minimizer itself. Considering the increasing interest in optimization by smoothing in recent years, these results may initiate a substrate for further theoretical studies of this method.

There are at least two important directions that can be pursued for further research. While our analysis was focused on smoothing by the Gaussian kernel, such kernel is not the only choice for smoothing. In fact, there is a rich literature about other smoothing kernels and their theoretical properties, such as Poisson kernels [60], Bessel kernels [61], etc. It would be interesting to explore the asymptotic behavior of smoothing by these other kernels.

Another direction for future research is seeking additional properties of a function which guarantees a traceable path from the asymptotic minimizer to some minimizer in the original (non-smoothed) function. More precisely, the Hessian of $g(\boldsymbol{x};\sigma)$ should not become singular along the followed path in order to ensure traceability of the minimizer. We believe, some constraints on the smoothness of f may provide control over the evolution of the eigenvalues of $\nabla^2 g(\boldsymbol{x};\sigma)$ over time.

6.2 Kernels

This dissertation studied the problem of signal blurring for the purpose of alignment by direct intensity-based methods. We argued that the use of traditional Gaussian image blurring, mainly inspired by the work of Lucas and Kanade [36], may not be suitable for non-displacement motions. Instead, we suggested directly smoothing the alignment objective function. This led to a rigorous derivation of spatially varying kernels required for smoothing the objective function of common model-based alignment tasks including affine and homography models.

The derivation process of the kernels in this dissertation may provide some insights for blur kernels in other tasks such as image deblurring, motion from blur, matching, optical flow, etc. For example, in image deblurring, the blur caused by the motion of the camera or by scene motion typically leads to spatially varying blur. The estimation of such kernels is very challenging [62–64]. Yet if the motion is close to the models discussed in this dissertation, our results may provide some insights for estimation of the blur kernel. Similarly, our kernels could be relevant to tasks involving motion blur [65], due to the physical relationship between motion estimation and blur estimation [66]. The coarse-to-fine scheme is a classic and very effective way to escape from poor local minima in optical flow estimation [67, 68]. Using the proposed kernels may boost the quality of the computed solution.

Another possible application which may benefit from our proposed kernels is visual detection and recognition. Heuristic spatially-varying kernels [26,27] have been successfully utilized in face detection [69] and object recognition [70,71]. Thus, our results may provide new perspective on using blur kernels for such tasks in a more principled way. Another related machinery for visual recognition tasks is convolutional deep architectures [72–76]. These methods apply learnable convolution filters to the scale-space representation of the images, hence gain translation and scale invariance. Utilizing the proposed kernels instead of traditional convolutional filters and scale-space representation between layers might extend the invariance of these methods to a broader range of transformations.

Finally there is a lot of room to improve the computational efficiency of using the proposed kernels. In this work, the integral transforms are evaluated on a dense grid. However, since the kernels are smooth and localized in

space, one might be able to get a good approximate of the integral transform by merely evaluating it at a small subset of image points.

REFERENCES

[1] K. Cordes, B. Rosenhahn, and J. Ostermann, "Increasing the accuracy of feature evaluation benchmarks using differential evolution," in *IEEE Symposium Series on Computational Intelligence (SSCI) - IEEE Symposium on Differential Evolution (SDE)*, apr 2011.

[2] D. Lowe, "Object recognition from local scale-invariant features," 1999.

[3] S. Vavasis, *Nonlinear optimization: complexity issues*, ser. International series of monographs on computer science. Oxford University Press, 1991. [Online]. Available: http://books.google.com/books?id=AdlQAAAAMAAJ

[4] P. J. M. van Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications*, 1987.

[5] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM J. on Optimization*, vol. 11, no. 3, pp. 796–817, Mar. 2000. [Online]. Available: http://dx.doi.org/10.1137/S1052623400366802

[6] C. A. Floudas, *Deterministic Global Optimization: Theory, Methods and (NONCONVEX OPTIMIZATION AND ITS APPLICATIONS Volume 37) (Nonconvex Optimization and Its Applications)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005.

[7] A. Morgan, *Solving Polynomial Systems Using Continuation for Engineering and Scientific Problems*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2009.

[8] A. Sommese and C. Wampler, *The Numerical Solution of Systems of Polynomials Arising in Engineering and Science*. World Scientific, 2005. [Online]. Available: http://books.google.com/books?id=S6fIrWaFN0sC

[9] A. Blake and A. Zisserman, *Visual Reconstruction*. MIT Press, 1987.

[10] L. T. Watson, "Theory of globally convergent probability-one homotopies for nonlinear programming." *SIAM Journal on Optimization*, pp. 761–780, 2001.

[11] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," in *Proceedings of the IEEE*, 1998, pp. 2210–2239.

[12] L. Piela, J. Kostrowicki, and H. A. Scheraga, "On the multiple-minima problem in the conformational analysis of molecules: deformation of the potential energy hypersurface by the diffusion equation method," *Journal of Physical Chemistry*, vol. 93, no. 8, pp. 3339–3346, 1989.

[13] V. Sindhwani, S. S. Keerthi, and O. Chapelle, "Deterministic annealing for semi-supervised kernel machines," in *ICML '06*. New York, NY, USA: ACM, 2006, pp. 841–848.

[14] P. Gehler and O. Chapelle, "Deterministic annealing for multiple-instance learning," in *AISTATS 2007*. Brookline, MA, USA: Microtome, 3 2007, pp. 123–130.

[15] M. Kim and F. D. Torre, "Gaussian processes multiple instance learning," pp. 535–542, 2010.

[16] P. S. Dhillon, S. S. Keerthi, K. Bellare, O. Chapelle, and S. Sundararajan, "Deterministic annealing for semi-supervised structured output learning." in *AISTATS 2012*, vol. 15, 2012.

[17] U. D. Hanebeck, K. Briechle, and A. Rauh, "Progressive bayes: A new framework for nonlinear state estimation," in *Proceedings of SPIE, AeroSense Symposium*, ser. SPIE '03, 2003, pp. 256–267.

[18] S. Schelstraete, W. Schepens, and H. Verschelde, "Energy minimization by smoothing techniques : a survey." pp. 129–185, 1999.

[19] Y. Bengio, *Learning Deep Architectures for AI*. Now Publishers Inc., 2009.

[20] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *International Conference on Machine Learning, ICML*, 2009.

[21] M. P. Kumar, B. Packer, and D. Koller, "Self-paced learning for latent variable models." in *NIPS*. Curran Associates, Inc., 2010, pp. 1189–1197.

[22] V. I. Spitkovsky, H. Alshawi, and D. Jurafsky, "From Baby Steps to Leapfrog: How "Less is More" in unsupervised dependency parsing," in *Proc. of NAACL-HLT*, 2010.

[23] I. Diener, "Trajectory methods in global optimization," in *Handbook of Global Optimization*. Kluwer Academic, 1995, pp. 649–668.

[24] A. C. Sun and W. D. Seider, "Recent advances in global optimization," C. A. Floudas and P. M. Pardalos, Eds. Princeton, NJ, USA: Princeton University Press, 1992, ch. Homotopy-continuation algorithm for global optimization, pp. 561–592. [Online]. Available: http://dl.acm.org/citation.cfm?id=137305.137347

[25] M. Loog, J. J. Duistermaat, and L. Florack, "On the behavior of spatial critical points under gaussian blurring. a folklore theorem and scale-space constraints," in *3rd Int. Conf. on Scale-Space and Morphology in Computer Vision*. London, UK: Springer-Verlag, 2001, pp. 183–192.

[26] A. C. Berg and J. Malik, "Geometric blur for template matching," in *2001 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2001, pp. 607–614.

[27] E. Tola, V. Lepetit, and P. Fua, "A fast local descriptor for dense matching," in *Conference on Computer Vision and Pattern Recognition*, Alaska, USA, 2008.

[28] D. V. Widder, *The Heat Equation*. Academic Press, 1975.

[29] M. Irani and P. Anandan, "All about direct methods," in *ICCV workshop on Vision Algorithms*, ser. ICCV '99. IEEE Computer Society, 1999, pp. 267–277.

[30] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the International Conference on Computer Vision*, ser. ICCV '99. IEEE Computer Society, 1999, pp. 1150–1157.

[31] H. Mobahi, Z. Zhou, A. Y. Yang, and Y. Ma, "Holistic 3d reconstruction of urban structures from low-rank textures," in *3DRR Workshop*, ser. ICCV'11, 2011, pp. 593–600.

[32] R. Szeliski, "Image alignment and stitching: a tutorial," *Found. Trends. Comput. Graph. Vis.*, vol. 2, pp. 1–104, January 2006.

[33] M. Cox, S. Lucey, and S. Sridharan, "Unsupervised alignment of image ensembles," Auto. Sys. Lab, CSIRO ICT Centre, Tech. Rep. CI2CV-MC-20100609, 2010.

[34] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework: Part 1," *International Journal of Computer Vision*, vol. 56, pp. 221–255, 2004.

[35] J. Jiang, S. Zheng, A. W. Toga, and Z. Tu, "Learning based coarse-to-fine image registration," pp. 1–7, 2008.

[36] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *IJCAI '81*, April 1981, pp. 674–679.

[37] P. Simard, Y. LeCun, and J. S. Denker, "Efficient pattern recognition using a new transformation distance," in *NIPS' 93*, San Francisco, CA, USA, 1993, pp. 50–58.

[38] F. d. l. Torre and M. J. Black, "Robust parameterized component analysis," ser. ECCV'02, 2002, pp. 653–669.

[39] N. Vasconcelos and A. Lippman, "Multiresolution tangent distance for affine-invariant classification," *NIPS' 98*, pp. 843–849, 1998.

[40] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Comm. of the ACM*, vol. 24, no. 8, pp. 381–395, 1989.

[41] H. Mobahi, Y. Ma, and L. Zitnick, "Seeing through the Blur," in *Proceedings of CVPR 2012*, 2012.

[42] M. Lefebure and L. D. Cohen, "Image registration, optical flow and local rigidity," *Journal of Mathematical Imaging and Vision*, vol. 14, pp. 131–147, March 2001.

[43] D. Marr and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London Series B*, vol. 207, pp. 187–217, 1980.

[44] J. J. Koenderink, "The structure of images," *Biological Cybernetics*, vol. 50, no. 5, pp. 363–370–370, Aug. 1984.

[45] A. P. Witkin, "Scale-Space Filtering." in *8th Int. Joint Conf. Artificial Intelligence*, vol. 2, Karlsruhe, Aug. 1983, pp. 1019–1022.

[46] A. L. Yuille and T. A. Poggio, "Scaling theorems for zero crossings," *IEEE PAMI*, vol. 8, no. 1, pp. 15–25, 1986.

[47] T. Lindeberg, "On the Axiomatic Foundations of Linear Scale-Space: Combining Semi-Group Structure with Causality vs. Scale Invariance," in *Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory*, 1994.

[48] T. Lindeberg and J. Garding, "Shape-adapted smoothing in estimation of 3-d depth cues from affine distortions of local 2-d brightness structure," in *Image and Vision Computing*, 1994, pp. 389–400.

[49] T. Lindeberg, "Generalized gaussian scale-space axiomatics comprising linear scale-space, affine scale-space and spatio-temporal scale-space," *J. Math. Imaging Vis.*, vol. 40, pp. 36–81, 2011.

[50] G. Osterberg, *Topography of the layer of rods and cones in the human retina*, ser. Acta ophthalmologica: Supplementum. Levin & Munksgaard, 1935.

[51] L. Sevilla and E. Learned-Miller, "Distribution fields," Dept. of Computer Science, University of Massachusetts Amherst, Tech. Rep. UM-CS-2011-027, 2011.

[52] D. Zwillinger, *Handbook of Integration*, ser. Ak Peters Series. Jones and Bartlett, 1992.

[53] S. Winitzki, ""a handy approximation for the error function and its inverse"," 2008.

[54] N. Snavely, S. Seitz, and R. Szeliski, "Modeling the world from internet photo collections," vol. 80, pp. 189–210, 2008.

[55] P. Torr, "Geometric motion segmentation and model selection," *Philosophical Transactions of the Royal Society of London*, vol. 356, no. 1740, pp. 1321–1340, 1998.

[56] I. Esteban, J. Dijk, and F. Groen, "Fit3d toolbox: multiple view geometry and 3d reconstruction for matlab." in *International Symposium on Security and Defence Europe (SPIE)*, 2010.

[57] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, 1992.

[58] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ser. SIGGRAPH '96. New York, NY, USA: ACM, 1996. [Online]. Available: http://doi.acm.org/10.1145/237170.237269 pp. 303–312.

[59] G. Turk and M. Levoy, "Zippered polygon meshes from range images," in *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, ser. SIGGRAPH '94. New York, NY, USA: ACM, 1994. [Online]. Available: http://doi.acm.org/10.1145/192161.192241 pp. 311–318.

[60] M. Felsberg and G. Sommer, *Scale Adaptive Filtering Derived from the Laplace Equation*, ser. Lecture Notes in Computer Science, Volume 2191. Springer Berlin Heidelberg, 2001, pp. 124–131.

[61] B. Burgeth, S. Didas, and J. Weickert, "The bessel scale-space," in *DSSCV*, ser. Lecture Notes in Computer Science, vol. 3753. Springer, 2005, pp. 84–95.

[62] A. Chakrabarti, T. Zickler, and W. T. Freeman, "Analyzing spatially-varying blur," ser. CVPR'10, 2010.

[63] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, "Non-uniform deblurring for shaken images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[64] A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless, "Single image deblurring using motion density functions," in *ECCV*, 2010.

[65] T. S. Cho, A. Levin, F. Durand, and W. T. Freeman, "Motion blur removal with orthogonal parabolic exposures," in *Int. Conf. in Comp. Photography (ICCP)*, 2010.

[66] W.-G. Chen, N. Nandhakumar, and W. N. Martin, "Image motion estimation from motion smear-a new computational model," *IEEE PAMI*, vol. 18, no. 4, pp. 412–425, 1996.

[67] L. Alvarez, J. Weickert, and J. Sánchez, "A scale-space approach to nonlocal optical flow calculations," ser. Scale-Space'99, 1999, pp. 235–246.

[68] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers, "Statistical and geometrical approaches to visual motion analysis," 2009, ch. An Improved Algorithm for TV-L1 Optical Flow, pp. 23–45.

[69] T. L. Berg, A. C. Berg, J. Edwards, M. Maire, R. White, Y.-W. Teh, E. Learned-Miller, and D. A. Forsyth, "Names and faces in the news," ser. CVPR, 2004.

[70] A. C. Berg, T. L. Berg, and J. Malik, "Shape matching and object recognition using low distortion correspondences," in *Proceedings of CVPR 2005.*, vol. 1, 2005, pp. 26–33.

[71] A. Frome, F. Sha, Y. Singer, and J. Malik, "Learning globally-consistent local distance functions for shape-based image retrieval and classification," in *In ICCV*, 2007.

[72] Y. Lecun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Int. Symp. on Circuits and Systems (ISCAS'10)*, 2010.

[73] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *ICML'09*, 2009, pp. 609–616.

[74] H. Mobahi, R. Collobert, and J. Weston, "Deep learning from temporal coherence in video," ser. ICML'09, 2009, pp. 737–744.

[75] M. Ranzato, J. Susskind, V. Mnih, and G. Hinton, "On deep generative models with applications to recognition," in *CVPR'11*, 2011.

[76] M. D. Zeiler, D. Kirshnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *CVPR'10*, 2010.